Bacterial community structure and composition of tropical river and drinking water – insights from deep sequencing and correlation to environmental drivers

Von der Fakultät für Lebenswissenschaften

der Technischen Universität Carolo-Wilhelmina zu Braunschweig

zur Erlangung des Grades

eines Doktors der Naturwissenschaften

(Dr. rer. nat.)

genehmigte

D i s s e r t a t i o n

von Douglas Mushi

aus Kilimanjaro, Tanzania

1. Referent:	apl. Professor Dr. Manfred G. Höfle
2. Referent:	Professor Dr. Michael Steinert
eingereicht am:	05.10.2015
mündliche Prüfung (Disputation) am:	07.12.2015

Druckjahr 2015

Vorveröffentlichungen der Dissertation

Teilergebnisse aus dieser Arbeit wurden mit Genehmigung der Fakultät für Lebenswissenschaften, vertreten durch den Mentor der Arbeit, in folgenden Beiträgen vorab veröffentlicht:

Tagungsbeiträge

Mushi D, Pereira RPA, Höfle, MG and Brettar I: Bacterial communities and their pathogenic potential along a pollution gradient in a tropical watershed in Tanzania. (Poster) PS-S1.12. *13th SAME conference on aquatic microbial ecology, EMBO, 8-13 September 2013, Stresa, Italy.*

Mushi D, Pereira RPA, Höfle, MG and Brettar I: Deep sequencing of microbial communities in tropical drinking water treatment plant reveals significant amplification of rare taxa. (Talk) *18th International symposium on health related water microbiology*, *WaterMicro2015*, *13-19 September, 2015, Lisbon, Portugal*.

The only true wisdom is in knowing you know nothing

Socrates (469-369 BC)

Acknowledgements

I wish to thank, first and foremost, the Tanzanian Ministry of Education and Vocational Training (MoEVT), Deutscher Akademischer Austauschdienst (DAAD, Funding programme number 50015773), and AQUAVALENS EC project (Grant agreement number 311846) for financing this research project.

The opportunity to work with microbial diagnostic (MIDI) research group under the guidance of Prof. Dr. Manfred G. Höfle and Dr. Ingrid Brettar that has significantly enhanced my knowledge in molecular diagnostics is substantially acknowledged.

It gives me great pleasure in acknowledging the thesis committee members, Prof. Carlos A. Guzman, Prof. Manfred G. Höfle, Prof. Andreas H. Farnleitner and Dr. Ingrid Brettar for their constructive criticism and invaluable advice that led to the successful completion of this work.

Prof. Dr. Michael Sternert is highly thanked for kindly accepting the offer for the second review of this thesis.

During the course of this research project, the support from my colleagues in the MIDI research group has been always congenial. Acknowledgement goes to, René Lesnik, Marina Pecellin, Verena Maiberg, Ashraf Zayed for their assistances in various stages of this research project, and Josefin Koch for introducing me to various lab facilities when I started working in the MIDI research group. Special thanks go to Rui Pereira for his invaluable advice and tireless assistance during library preparation, deep sequencing and data analysis.

A special note of thanks to Prof. Dr. Allen Lewis Malisa, Head, Department of Biological Sciences, Sokoine University of Agriculture, who made the laboratory available for me to process the water samples, and to Mr. James Mwesongo, for his diligent technical assistance during field work in Morogoro, Tanzania.

I wish to express my heartfelt gratitude to Michael Böcher, Gang Zhao, Hui Wang and Clemens Schmidt whose cooperation, solidarity and understanding made my stay in Germany comfortable and above all unforgettable.

Last but not least, I would like to thank my parents Mr. and Mrs. Werasimbo A. Mushi for allowing me to realize my own potential. The immense support they have provided me over a long period of time was the distinctive gift anybody has ever given me, ``**asanteni sana**''.

Table of Contents

Abbreviations

CP	*Clostridium perfringens*
TC	Total coliforms
HPC	Heterotrophic plate counts
BOD	Five day's biological oxygen demand
TSS	Total suspended solids
DO	Dissolved oxygen
CV	Coefficient of variation
ANOSIM	Analysis of similarity
SIMPROF	Similarity profiles
OTUs	Operational taxonomic units
RDP	Ribosomal database project
DWTP	Drinking water treatment plant
PCA	Principal component analysis

Summary

Tropical river systems are critical resources to human civilization, and for maintenance of ecological services. Lack of protection exposes river systems to natural and anthropogenic influences. The consequences of these influences do not only shape riverine critical ecological services but also significantly increase cost of treating river water for drinking purposes, limit water supply, and may put human populations to unacceptable risks. Reliable tools for detecting anthropogenic influences occurring in tropical river systems have not been comprehensively evaluated. As a consequence, major contamination sources in tropical river systems remain elusive. Furthermore, direct application of detection tools applied in other regions such as temperate for riverine monitoring may not yield realistic information because of substantial variations in basin characteristics (such as climate, hydrology, natural vegetation), mineral composition of sediment/bedrock, type and degree of anthropogenic influences between rivers of different climatic conditions.

In this study, 17 quantitative physico-chemical, microbiological and trophic level variables were determined in a tropical riverine system with a watershed dominated by pristine forest, urban, and agricultural land use types during low and high hydrological regimes. Principal component analysis showed that land use was the most influential factor driving the tropical riverine system. This finding was supported by the fact that significant correlations among physico-chemical and microbiological variables were only possible with pooled data of all land use types rather than individual land use data. Of all variables tested only biological oxygen demand, nutrients (nitrate and phosphate), *Clostridium perfringens* and chlorophyll-*a* significantly discriminated sampling sites according to watershed land use types indicating their high sensitivity towards diverse influences and demonstrated potential for future applications in management and mitigation measures of tropical river systems.

Characterisation of water samples collected on the basis of dominant watershed land-use types along the tropical riverine system using Illumina MiSeq sequencing of the 16S rRNA gene showed that riverine microbial communities were greatly discriminated by watershed land-use

types rather than environmental or climate characteristics. Watershed land-use promoted allochtonous bacterial sequences, eutrophication-linked bacterial sequences, extinction and stress to indigenous taxa. Despite the considerable variation across watershed land-use types, the overall microbial richness was primarily due to among-site richness than within-site richness. Microbial communities showed a remarkable stability over time irrespective of climatic perturbations emphasizing the possibility of being temporally predictable.

Comprehensive investigation of a tropical drinking water treatment plant fed by water from the pristine location of the investigated tropical riverine system showed coherent dynamics of taxa and microbial community shifts along the treatment barriers of the drinking water treatment plant. By sequencing 16S rRNA gene amplicons at adequate depth, a high degree of microbial diversity and overrepresentation of typical freshwater genera including *Undibacterium, Novosphingobium and Cylindrospermopsis* were observed. *Undibacterium* had a considerable contribution to the abundance of the phylum *Proteobacteria* and demostrated a remarkable ability to predict microbial diversity. Shifts in community structure were due to substantial elimination of bacterial taxa by sand filtration, and significant enrichment of rare abundant taxa following chlorination. Taxa coherent dynamics across treatment barriers revealed the presence of a series of discrete microbial secondary successions punctuated by treatment barriers. Based on microbial community succession data, the fate of noxious bacteria in drinking water treatment plant is potentially predictable.

The outcome of this study provided critical insights into the potential factors shaping the riverine system and drinking water treatment plant in tropical environment. The fact that this is the first comprehensive study covering the critical aspects of source and drinking water in tropical environment, the results can be of great importance in designing appropriate riverine and treatment plant management approaches, and identifying suitable monitoring tools in order to boost source water quality and performance of drinking water treatment plants in tropical environments similar to that of Morogoro in Tanzania.

CHAPTER 1

General Introduction

1.0 Riverine systems and drinking water supply

Riverine systems are critical natural resources serving as drinking water sources for many human societies world-wide (1–6), but they are also used for recreational, agricultural, hydropower, industrial and transport purposes (7–10). These services are continuously offered as far as the riverine integrity across spatial scales is adequately maintained. Unfortunately, riverine systems are directly impaired by human-derived activities, anthropogenic climate change, and a complex suit of natural processes (9,11–14). While riverine systems are controlled more by anthropogenic forcing than natural drivers (15), population growth, rapid economic growth and urbanization highly accelerate these anthropogenic influences (16). Commonly, the anthropogenic contaminants get into riverine water through erosion, run-off following rainfall, atmospheric deposition, and point sources such as wastewater discharge (11). Of note is that acute stresses exerted are diverse and result from many human activities such as heavy burden of pollution from residential, urban development, agricultural and industrial activities (Fig. 1), which pose risks to the quality and cost of drinking water and the reliability of water supplies (16).

The contaminations can pose significant threats to riverine ecosystems and drinking water resources (17). Erosion events tend to elevate levels of turbidity in river waters which cause deterioration in treatment performance and direct impact on coagulant demand during water treatment (18). Influxes of nutrients into the river system cause large algae blooms that generate negative aesthetic and eutrophic conditions (19) but also may cause elevated cyanobacterial toxins in river water, and demand extra treatment steps such as ozonation to remove cyanotoxins during drinking water treatments (11). Previous studies showed that contamination reduces water supply and increase drinking water treatment cost (16). Shannon *et al*., (20) found that river water contaminated with ammonia and organic nitrogen can

seriously reduce the efficacy of disinfection while elevated organic matter in the source water can lead to high levels of toxic disinfection by-products. Of great concern is the occurrence of waterborne pathogens which spread within the riverine and drinking waters after contamination by animal or human waste (11) and treatment failure, respectively leading to the increased probability of outbreaks of waterborne diseases.

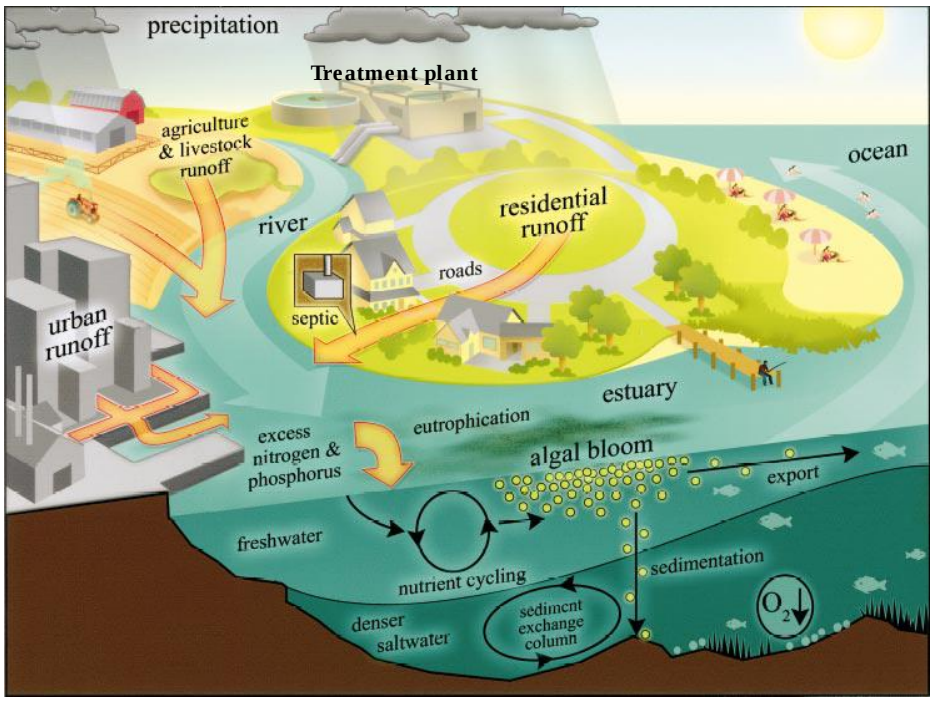


Figure 1: Sketch diagram showing a drinking water treatment plant and major human activities leading to serious threats in riverine systems. Adapted from Pearl *et al.*, (21).

Despite the fact that riverine systems are highly susceptible to contamination and waterborne pathogens as a consequence of in-adequate protection (2), many river systems in tropical environments are poorly studied regardless of being heavily exploited as the source of drinking water. Absence of frequent water quality monitoring programmes combined with the unique tropical climate that support diverse varieties of waterborne pathogens (22) suggests that many tropical river systems contain water of unknown quality, which may put people at an unacceptable risk of infection by pathogens (23) and toxic substances. Furthermore, tropical basin characteristics (such as climate, hydrology, natural vegetation), mineral composition of sediment/bedrocks (24), type and degree of anthropogenic influences do not allow direct

application of river monitoring tools developed in temperate environments without being thoroughly evaluated for their efficacy in river water quality monitoring of tropics.

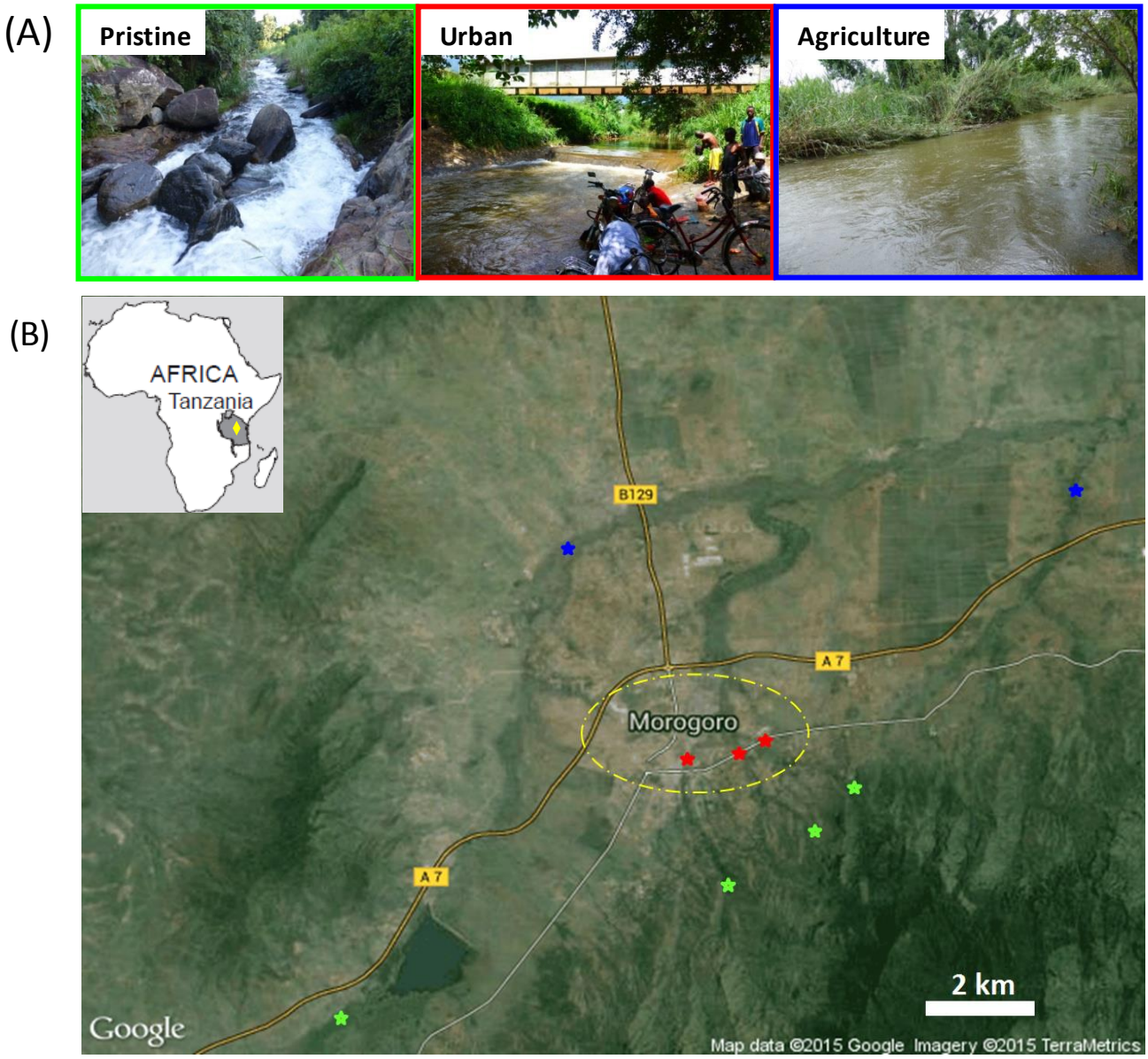


Figure 2: Geographic location of the investigated tropical river system. The watershed land-use scale consisted of three rank-ordered levels i.e. pristine forest, urban and agriculture. Informal views of representative sampling points in each land use type are portrayed (**A**). The satellite photograph (**B**) shows the distribution of sampling points in pristine forest (represented by green stars), urban (red stars), and agricultural (blue stars) land use types, respectively. The inset is the political map of Africa with Tanzania displayed in grey and the approximate location of the Morogoro River represented by a yellow diamond. The location of the Morogoro City is shown by a dashed yellow sphere.

The Morogoro River, located in a tropical East African country, Tanzania, serves as the main source of drinking water to more than 300,000 people residing in Morogoro City and to over 6 million in Dar es Salaam City, Tanzania. This river system drains extensively from a watershed with diverse land-use types including native forest, urban, and agricultural landscapes (Fig. 2). While native forest land-use is protected and devoid of human activities, the urban landscape is characterized by anthropogenically associated activities (e.g. municipal and industrial discharge, washing activities), and the agricultural landscape subjugated by grazing livestock (cattle, sheep, goat). Each of these land-use types differentially influences the riverine water system. However sensitive monitoring tools appropriate for identifying these types of anthropogenic influences and discriminating them from those linked to natural processes are unknown for the Morogoro Riverine system despite being heavily exploited as a source of water supply (25). This impairs effective management of river health and limits our understanding of the interaction between anthropogenic and riverine bacterioplankton. Consequently, the necessity to perform a comprehensive study in the Morogoro River system to identify the sensitive and reliable monitoring tool(s) to guarantee effective and efficient tropical riverine management is thus apparent (14).

Like other surface waters, the Morogoro River system carries a considerable load of dissolved and particulate materials from both natural and anthropogenic sources downstream (14). Consequently, this river water requires thorough treatments that can reduce chemical substances, inactivate noxious microbiota and limit their re-growth before being distributed to the end users. For decades, conventional drinking water treatment plants (DWTP) equipped with a train of treatment processes (e.g. aeration, coagulation/flocculation, sedimentation, sand filtration and post-chlorination) have been operating in Tanzania to treat river water for domestic and industrial purposes. Despite the frequency of occurrence of waterborne disease outbreaks associated with the drinking water (26), the impact of each of these treatment steps on the quality of finished water in general, and on microbial communities in particular, is basically unknown for a tropical DWTP such as that of Morogoro City in Tanzania.

The following critical questions were addressed in order to get insights into the responsiveness of tropical river system against watershed land use influences, and the performance of a drinking water treatment plant fed by the Morogoro River.

(1) What quantitative environmental and microbiological water column variables considerably and differently respond to influences emanating from distinct land-use types dominating the tropical riverine watershed? Are these variables able to identify influential sources of contamination in tropical riverine system?

(2) What determines the bacterioplankton community composition and structure in a tropical riverine system?

(3) Does the microbial community structure and composition allow an understanding of the performance of a drinking water treatment plant treating water from a tropical river system?

In this study, the impact of watershed land use on the ecological health of Morogoro River system situated in a tropical East African environment (Tanzania) was assessed. This included the responsiveness of 13 quantitative water column physico-chemical (pH, dissolved oxygen, biological oxygen demand, conductivity, ammonia, nitrate, nitrite, phosphate, chloride, total hardness), microbiological (*Clostridium perfringens*, total coliforms, heterotrophic plate counts, total bacterial cell counts) and trophic level (Chlorophyll-*a*) parameters to pristine forest, urban and agricultural land use influences. The fluctuation of each variable within and between land use types across low and high hydrological regimes was comprehensively analysed in order to understand in detail the capacity of each variable in detecting extrinsic influences under distinct situations. Applications of multivariate analysis allowed identification of influential sources of contamination in the studied tropical riverine system. Breadth of contamination was assessed by comparing values of each variable detected in pristine forests against those observed in urban or agricultural land uses. It should be noted that the characteristics of the pristine forest

location equals those described by Sanchez-Montoya *et al*., (24) for the reference sites making the determination of contamination breadth possible.

Microbial community structure and composition of the investigated tropical river system were determined by deep sequencing of the 16S rRNA gene spanning V4 and V5 hypervariable regions using an Illumina MiSeq platform. Comparison of 16S rRNA data set between land use types allowed the assessment of influential factor driving the presence, structure and diversity of microbiota in a tropical riverine system.

The performance of tropical drinking water treatment plant was assessed by comparing microbial community profiles after each treatment barrier with that from the source water (Morogoro River water). The critical treatment steps altering the microbial community structure in tropical drinking water treatment plant were determined. Incidence of microbial succession in drinking water was assessed by the aid of index of associations. Re-growth incidence of rare taxa could be tracked by comparing microbial community profiles of source water with that from post-chlorination.

1.1 Literature cited

1. Hellard ME, Sinclair MI, Forbes AB, Fairley CK. A randomized, blinded, controlled trial investigating the gastrointestinal health effects of drinking water quality. Environ Health Perspect. 2001;109(8):773–8.
2. Kristemann T, Claßen T, Koch C, Dangendorf F, Fischeder R, Gebel J, Vacata V, et al. Microbial load of drinking water reservoir tributaries during extreme rainfall and runoff. Appl Environ Microbiol. 2002;68(5):2188–97.
3. Kosaka K, Asami M, Matsuoka Y, Kamoshita M, Kunikane S. Occurrence of perchlorate in drinking water sources of metropolitan area in Japan. Water Res. 2007;41(15):3474–82.
4. Reweta WSJ, Sampath RK. Performance Evaluation of urban water supply in Tanzania: The Case of Dar Es Salaam City. Int J Water Resour Dev. 2000;16(3):407–21.
5. Marobhe NJ. Critical review of water supply services in urban and rural areas of Tanzania. Water Policy 2008;10:57–71.
6. Komakech HC, Van der Zaag P, Van Koppen B. The last will be first: Water transfers from agriculture to Cities in the Pangani River Basin, Tanzania. Water Altern. 2012;5(3):700–20.

7. Staley C, Gould TJ, Wang P, Phillips J, Cotner JB, Sadowsky MJ. Bacterial community structure is indicative of chemical inputs in the Upper Mississippi River. Front. Microbiol. 2014; 5:524.
8. Savio D, Sinclair L, Ijaz UZ, Parajka J, Reischer GH, Stadler P, et al. Bacterial diversity along a 2600 km river continuum. Environ Microbiol. 2015;doi:10.1111/1462-2920.12886(Epub ahead of print).
9. Kirschner AKT, Kavka GG, Velimirov B, Mach RL, Sommer R, Farnleitner AH. Microbiological water quality along the Danube River: Integrating data from two whole-river surveys and a transnational monitoring network. Water Res. 2009;43(15):3673–84.
10. Pinto U, Maheshwari BL. River health assessment in peri-urban landscapes: An application of multivariate analysis to identify the key variables. Water Res. 2011;45:3915–24.
11. Delpla I, Jung AV., Baures E, Clement M, Thomas O. Impacts of climate change on surface water quality in relation to drinking water production. Environ Int. 2009;35(8):1225–33.
12. Bu H, Tan X, Li S, Zhang Q. Temporal and spatial variations of water quality in the Jinshui River of the South Qinling Mts., China. Ecotoxicol Environ Saf. 2010;73(5):907–13.
13. Kazi TG, Arain MB, Jamali MK, Jalbani N, Afridi HI, Sarfraz RA, et al. Assessment of water quality of polluted lake using multivariate statistical techniques: A case study. Ecotoxicol Environ Saf. 2009;72(2):301–9.
14. Zhang Z, Tao F, Du J, Shi P, Yu D, Meng Y, et al. Surface water quality and its control in a river with intensive human impacts-a case study of the Xiangjiang River, China. J Environ Manage. 2010;91:2483–90.
15. Meybeck M. Global analysis of river systems: from Earth system controls to Anthropocene syndromes. Phil Trans R Soc Lond B. 2003;358:1935–55.
16. Postel SL, Thompson BH. Watershed protection: Capturing the benefits of nature's water supply services. Nat Resour Forum. 2005;29(2):98–108.
17. St Laurent J, Mazumder A. The influence of land-use composition on fecal contamination of riverine source water in southern British Columbia. Water Resour Res. 2012;48(W00M03):1–11.
18. Shin JY, Spinette RF, O'Melia CR. Stoichiometry of coagulation revisited. Environ Sci Technol. 2008;42(7):2582–9.
19. Weiss PT, Erickson a J, Gulliver JS. Cost and Pollutant Removal of Storm-Water Treatment Practices. J Water Resour Plan Manag. 2007;133(3):218–29.
20. Shannon MA, Bohn PW, Elimelech M, Georgiadis JG, Mariñas BJ, Mayes AM. Science and technology for water purification in the coming decades. Nature. 2008;452:301–10.
21. Paerl HW, Valdes LM, Peierls BL, Adolf JE, Harding LW. Anthropogenic and climatic influences on the eutrophication of large estuarine ecosystems. Limnol Oceanogr. 2006;51(1, part 2):448–62.
22. Guernier V, Hochberg ME, Guégan JF. Ecology drives the worldwide distribution of human diseases. PLoS Biol. 2004;2(6):0740–6.
23. Mushi D, Byamukama D, Kirschner AKT, Mach RL, Brunner K, Farnleitner AH. Sanitary inspection of wells using risk-of-contamination scoring indicates a high predictive ability

for bacterial faecal pollution in the peri-urban tropical lowlands of Dar es Salaam, Tanzania. J Water Health. 2012;10(2):236–43.
24. Sánchez-Montoya MDM, Arce MI, Vidal-Abarca MR, Suárez ML, Prat N, Gómez R. Establishing physico-chemical reference conditions in Mediterranean streams according to the European Water Framework Directive. Water Res. 2012;46:2257–69.
25. Jiwa SF, Mugula JK, Msangi MJ. Bacteriological quality of potable water sources supplying Morogoro municipality and its outskirts: a case study in Tanzania. Epidemiol Infect. 1991;107:479–84.
26. Wang H, Edwards M, Falkinham JO, Pruden A. Molecular survey of the occurrence of legionella spp., mycobacterium spp., pseudomonas aeruginosa, and amoeba hosts in two chloraminated drinking water distribution systems. Appl Environ Microbiol. 2012;78(17):6285–94.

CHAPTER 2

Influential drivers of tropical riverine system revealed by statistical analysis of the water column microbiological and physico-chemical characteristics

Mushi D[1, 2], Höfle MG[1] and Brettar I[1]

[1]Helmholtz Centre for Infection Research, Braunschweig (Germany)

[2]Sokoine University of Agriculture, P.O. Box 3038, Morogoro, Tanzania

2.0 Abstract

We surveyed a tropical riverine system consisting of diverse watershed characteristics including pristine forest, urban and agricultural land uses. Effect of each land use type on microbiological and physico-chemical characteristics of riverine water column was investigated during the low and high hydrological regimes. Land-use and natural processes were identified as the most influential drivers of river water quality by principal component analysis at a total of 85.2% explainable variations. Significant correlation among physico-chemical and microbiological variables were only possible with pooled data of all land use types rather than individual land use data indicating that anthropogenic gradient was well reflected. Watershed land use classification was only possible by specific individual variables highly reactive to land use influences justifying their use for future management and mitigation measures of tropical riverine system.

2.1 Introduction

The Morogoro River is an essential ecosystem serving as a source of water for domestic, agricultural, and industrial purposes in Morogoro City as well as Dar es Salaam City in Tanzania (1). As a consequence, maintenance of its microbiological and physico-chemical quality is of paramount importance for sustainable development. However, this river system originates from a protected environment with native forest vegetation and flows through human dominated urban and livestock grazing landscapes. Through various mechanisms acting over a range of spatial and temporal scales, each of these landscapes may contribute to microbiological and physico-chemical composition of river water (2–6). Unfortunately, we do not precisely know to what extent this alteration may affect the critical services provided by riverine system. On the other hand, microbiological and physico-chemical variables in river water column may have multiple sources including sediment, bedrocks, biogeochemical transformations, point and non-point sources (7–10). Consequently, there is a need to identify major source(s) impacting river water in order to precisely estimate the health risks associated

with tropical river water and take appropriate remedial actions for microbiological and physico-chemical quality improvements.

Worldwide, there exist an ample number of studies that have assessed water quality of various riverine systems from local to regional and transnational scales (11–13). However, their findings cannot be generalized to other riverine systems such as that of tropical environment due to substantial variations in their basin characteristics (such as climate, hydrology, natural vegetation), mineral composition of sediment and bedrocks (12), type and degree of anthropogenic influences. The Morogoro River has never been comprehensively studied in the light of microbiological and physico-chemical water quality despite serving as a substantial source of water for domestic and industrial purposes (1). Consequently, formulation of a sound programme appropriate for an effective water quality motoring will depend on sensitive and informative variables selected according to their responsiveness against major disturbances identified in the Morogoro Riverine system.

The current study assessed the influence of watershed land-use types on microbiological and physico-chemical characteristics of the Morogoro Riverine system located in a tropical Eastern African country, Tanzania. We selected sampling sites in headwaters dominated by pristine forest, and that fulfilled twenty criteria of reference sites described in detail by Sánchez-Montoya *et al*., (12,14). Sampling sites established in urban, and agricultural locations of the Morogoro Riverine system were selected following criteria established by Snelder and Biggs (15), however neither of these sites could fulfill criteria of reference sites. From several sites of each location, we determined multiple physico-chemical, bacteriological and trophic level variables spanning high and flow regimes of the river system. As expected, pristine sampling points showed substantially low values of bacteriological and physico-chemical quality and provided a basis for assessing the influence of agricultural and urban land uses in riverine water quality. Significant correlation among physico-chemical and microbiological variables were only possible with pooled data of all land use types rather than with data from individual land use indicating that presumed anthropogenic gradient were well reflected. Land-use was identified

as the main factor influencing river water quality. Specific individual variables showed high sensitivity to land-use influences and may be promising candidates for effective water quality monitoring and evaluation of riverine system under restoration.

2.2 Materials and Methods

Morogoro riverine system belongs to the upper Ngerengere watershed geographically located between 6°51´, 7°09´S and 37°32´, 38°38´E at an altitude ranging from 500 to 2260m above sea level (see Fig. 2B in chapter 1). The watershed topography ranged from mountainous in the North-Western Uluguru mountains (2,260 m a. s. l.)–where the river originates–to low land in the downstream reaches (16). The river extensively drains native forest of Uluguru Mountains, urban landscape at the foot of Uluguru Mountains, and livestock grazing land in the downstream reaches, respectively.

The tropical climate of the investigated watershed is characterized by ambient temperature, dropping below 20^{o}C during the period covering May and June. Annual rainfall ranges from 800 to more than 1,500 mm with a bimodal regime (17). More than seventy percent of precipitation drops between March and May, the time when the hydrographs of the river indicate peak flow, whereas the remaining percentage accounted for the small peak that happens between November and April. Despite serving as a source of drinking water for the Cities of Morogoro and Dar es Salaam in Tanzania, the discharge of Morogoro riverine system ranged from 500 l/s in the headwaters to >3000 l/s in the downstream reaches during the time of this survey.

A total of 90 water samples was taken from 9 synoptic sampling points established along the Morogoro Riverine system on the basis of adjoining watershed land-use types. These included four synoptic sites established at pristine forest location, three sites at urban location, and two sites at agricultural location, respectively. Watershed land-use types were delineated according to Snelder and Biggs (15). Each site was sampled 10 times during the entire study period that spanned low and high flow conditions.

Pristine forest landscape was located in the North-Western Uluguru Mountain within Uluguru nature reserve. While wild animals (such as vervet monkeys, wild pig, duiker antelopes, wild birds) and native forest dominated this location, neither human settlements nor anthropogenic activities were observed within and around this environment over the entire study period (see Fig. 2A in chapter 1). Headwaters in pristine forest flow over Precambrian rocks of meta-sedimentary nature and fulfilled all the 20 criteria for reference sites (see reference 12,14). Urban land-use adjoining the Morogoro Riverine system is located immediately after the foot of Uluguru Mountain. A population of >300,000 people (Tanzania National Bureau of Statistics) inhabits this landscape. At this landscape, river water flows over sediments of tertiary and quaternary ages. Domestic waste water and treated sewage from urban inhabitants and municipal sewage treatment plants, respectively contributed to riverine system. Additionally, washing activities could be discernible (see Fig. 2A in chapter 1). Ipso facto, the urban location could not fulfill any of the 20 criteria for reference sites. Further downstream, agricultural land-use (Fig. 2A in chapter 1) dominated the watershed that involved livestock grazing (mainly cattle, goats and sheep) with direct access to river water especially during watering. Organic matter leaking from the organic farming systems adjacent to the river water was also evident. As in urban, river water at agricultural watershed land-use streamed over sediments of tertiary and quaternary ages and did not satisfy any of the 20 criteria for reference sites.

Dissolved oxygen (DO), temperature, conductivity and pH were assayed *in situ* with PCE-PHD 1 meter (PCE Deutschland, Meschede, Germany). Probes were calibrated at 25 ^{o}C prior to sampling date and the calibration was substantiated after field measurements according to manufacturer´s instructions. Membrane electrode (4500-O G) and gravimetric techniques (2540D) were used to assay five day´s biochemical oxygen demand (BOD) and total suspended solids (TSS), respectively as described in American Public Health Association (APHA) standard methods for water quality analysis (APHA 2000). Semi-quantitative QUANTOFIX® test strips (Macherey-Nagel, Düren, Germany) were used to estimate the concentration of ammonium (NH_4), phosphate (PO_4), nitrate (NO_3), nitrite (NO_2), chloride (Cl) ions and hardness in the river water according to manufacturer's instructions. Water samples were taken aseptically from the

middle of the river–at a depth of approximately 30 cm–using 1 l wide-mouthed sterile plastic bottles (Thermo Scientific™ Nalgene™, Neubrecht, USA). Samples were immediately placed into a dark, ice-cooled (4°C) box and analyzed within 6 to 9 hours after collection of the first sample.

Manufacturer-based directions were used for media preparation and bacterial colonies inventory. Heterotrophic plate counts (HPC) were performed by spreading an appropriate volume (0.001 to 1 ml) of the sampled river water on R2A agar (Carl Roth GmbH, Karlsruhe, Germany) plates before incubating at 22°C for 72 h. Total coliforms (TC) and *Clostridium perfringens* (CP) were determined by membrane filtration technique using selective media, Endo agar (EA) for TC (Merk, Darmstadt, Germany) and Fluorocult-tryptose sulfite cycloserine (F-TSC) for CP (Merk, Darmstadt, Germany). 0.001 to 100 ml of water sample was filtered through 0.45-µm-pore-size and 47-mm-diameter cellulose nitrate membrane filters (Sartorius Stedim Biotech GmbH, Goettingen, Germany). To enumerate TC, filters containing the trapped cells (with size >0.45 µm) were placed on the EA plates, and incubated at 37°C for 24 h. CP were isolated by heating the water sample at 75°C for 15 min in a water bath prior to filtration. After filtration, the membrane filters were placed on F-TSC Agar plates and incubated at 44°C for 24 h while in an anaerobic jar containing the Anaerocult A anaerobic system (Merck, Darmstadt, Germany). To enumerate HPC, all the colonies of different color and morphology growing on R2A agar were counted as HPC. On EA plates, colonies with characteristic pink to rose red with metalic sheen (*E. coli*), pink and mucoid (*Klebsiella pneumoniae*, *Enterobacter aerogenes*), and colorless to pale pink (*Salmonella typhi*, *Shigella sonnei*, *Proteus vulgaris*) were counted as TC, whereas black colonies on F-TSC agar that fluoresce after cleaving 4-methylumbelliferyl-phosphate to 4-methylumbelliferone in the presence of UV-light (366 nm) were counted as CP. All colony counts were expressed as colony forming units (cfu) per ml.

Total bacterial cell counts (called hereafter bacterial abundance) were performed by fixing appropriate volume of water sample with 2% formaldehyde and staining the fixed samples with 3.76 µl of Sybr green I dye (Molecular Probes, Invitrogen) prior to 15 min incubation in the dark

at room temperature. After incubation, 5ml of the stained sample was filtered through 0.2-µm-pore-size and 25-mm-diameter filter (Whatman Anodisc). Filter containing the trapped cells was placed on the microscopic slide pre-treated with Citifluor before being incubated at 4°C for 20 min. After incubation, total cells from the slide were analyzed by epifluorescence microscopy using Sybr green I dye related filter system (with excitation of 488 nm and emission of 500 nm wavelengths). By using 63 × magnifications, ten pictures of known dimensions were obtained from each slide. Image J software was used to separately quantify the total cells from each picture.

Table 1. Mean (±SD) physico-chemical water quality characteristics. H-*test* represent Kruskal-Wallis (***P<0.001, **P<0.01, *P<0.5, not significant (ns): P>0.05); effect change equals to the percentage increase from pristine to agriculture or urban. Different letters among locations indicate those values that were significant different from each other according to post-hoc multiple comparisons determined by Tukey's test. Abbreviations: Wtemp = water temperature; DO = dissolved oxygen; TSS = total suspended solids; BOD = Five days' biological oxygen demand.

Location	Wtemp (^{o}C)	pH	DO (mgl^{-1})	TSS (mgl^{-1})	BOD (mgl^{-1})	NH_4^+ (mgl^{-1})	PO_4^- (mgl^{-1})	NO_3^- (mgl^{-1})	NO_2^- (mgl^{-1})	Conductivity (µS)	Hardness (mgl^{-1})	Chloride (mgl^{-1})
Pristine (n=40)	18(±0.2)^{a}	8.0(±0.08)	9.0(±0.2)	0.2(±0.03)^{a}	1.7(±0.2)^{a}	0.3(±0.01)^{a}	7.9(±0.8)^{a}	2.7(±0.3)^{a}	0.1(±0.003)^{a}	53(±2)^{a}	95(±2.5)^{a}	23(±3.6)^{a}
Urban (n=30)	25(±0.5)^{b}	7.6(±0.1)	9.6(±0.6)	0.4(±0.1)^{a}	4.1(±0.4)^{b}	20(±4.6)^{b}	19 (±3.2)^{b}	43(±15.4)^{b}	6.3(±2.5)^{b}	829(±105)^{b}	276(±20)^{b}	280(±31)^{b}
Agriculture (n=20)	27(±0.7)^{b}	7.4(±0.1)	6.6±(0.9)	1.4(±0.5)^{b}	6.4(±0.4)^{c}	39(±7.7)^{c}	33 (±5.4)^{c}	7.8(±1.2)^{a}	0.4(±0.1)^{a}	592 (±68)^{b}	250(±24)^{b}	193(±39)^{b}
H-*test*	0.01*	ns	ns	0.006**	0.001***	0.001***	0.001***	0.001***	0.001***	0.001***	0.001***	0.001***
Effect size	0.5	-0.08	-0.3	6.0	2.8	194	3.2	14.9	62.0	14.6	1.91	11.2

Table 2. Microbiological (log+1 transformed) water quality and trophic level characteristics. Abbreviations: CP = *Clostridium perfringens*; TC = Total coliforms; HPC = Heterotrophic plate counts; CV = non parametric coefficient of variations (CV = (75 Quantile - 25 Quantile)/50 Quantile); Bacterial abundance were microscopically determined. For details of H-*test*, different letters among locations and effect size see caption in Table 1.

Location	Bacteriological counts (log cfu or log cells per ml)												Chlorophyll-*a* (µg/l)		
	CP			TC			HPC			Bacterial abundance					
	Median	Range	CV	Median	Range	CV	Median	Range	CV	Median	Range	CV	Median	Range	CV
Pristine (n=40)	0.3[a]	0.04 - 0.9	1.6	2.3[a]	1.9 - 2.5	0.1	3.6[a]	3.1 - 4.0	0.1	6.1[a]	5.9 - 6.3	0.02	0.3[a]	0.2-0.8	0.5
Urban (n=30)	1.3[b]	1.0 - 2.7	1.2	4.1[b]	3.0 - 4.8	0.2	4.8[b]	4.3 - 5.8	0.2	7.5[b]	7.2 - 8.8	0.04	5.0[b]	1.6 - 9.6	1.2
Agriculture (n=20)	2.1[c]	1.1 - 2.9	1.0	3.6[b]	3.1 - 4.0	0.4	5.4[b]	5.0 - 6.0	0.4	8.4[b]	7.3 - 8.7	0.3	3.0[c]	1.1 - 7.7	1.3
H-*test*	0.001**			0.032*			0.041*			0.036*			0.001**		
Effect size	6.0			0.8			0.5			0.38			15.7		

Chlorophyll-*a* concentration ($\mu g l^{-1}$) was determined spectrophotometrically following pigment extraction with 94% ethanol. Briefly, 200 to 3000 ml of water samples was filtered through a glass fiber filter (Whatman GF/F, 47 mm diameter, Maidstone, UK) on the day of sampling. After filtration, the filter was placed in to an extraction tube containing 20 ml of 94% ethanol. The mixture was shaken strongly prior to dark incubation at 4 ^{o}C overnight. After incubtion, the extract was filtered using a glass fiber filter (Whatman GF/F, 25 mm diameter, Maidstone, UK). Ten ml of clear extract were acidified by 10 µl of 3 mol l^{-1} HCl. The absorbance of clear extracts with and without acid was measured at 665 and 750 nm. The concentration of chlorophyll-*a* was determined according to Kirschner *et al*., (11).

Our investigational design involved a watershed land-use scale consisting of three rank-ordered levels i.e. pristine forest, urban and agricultural. We tested the potential connection between land-use and water column microbiological and physicochemical characteristics. Non-parametric, rank based H-*test* (Kruskal-Wallis ANOVAs) was used to test for effects of watershed land-use on physico-chemical parameters. H-*test* was employed due to reasonable sample size gathered and distributional characteristics of the data (2). Post-hoc pair-wise comparisons were performed by Tukey's test once H-*test* showed significant difference between watershed land-uses for a given parameter (2). To test the breadth of land-use effect on various response variables, a unit-free effect size (2) was computed as the percentages of increase from the reference pristine to urban or agricultural extreme mean/median value. Fulfillment of data to Kaiser-Meyer-Olkin's (KMO) condition permitted the use of principal component analysis (PCA) to define the possible contributing factors to riverine water quality deterioration. Sites clustering was performed using the complete linkage method based on coefficients of similarity measured by Euclidean distance (similar clusters were generated by methods like group average or single linkage) and physico-chemical or microbiological data normalized to log+1. We used spearman rank correlation to determine variable associations, and to corroborate PCA results. All statistics were performed using SPSS for Windows, version 16.0 (SPSS Inc., Chicago, Illinois).

Table 3: Principal component analysis (PCA) extracted from microbiological and physico-chemical data set generated from 90 river water samples. Note that correlation coefficients that were not significant are presented as `ns´. For abbreviations see captions of Table 1 and 2, respectively.

	Correlation coefficients	
Variable	PC1	PC2
Temperature	0.90	ns
pH	ns	0.85
Conductivity	0.71	0.63
Dissolved oxygen	-0.56	0.71
TSS	0.87	ns
BOD	0.98	ns
Ammonium	0.96	ns
Phosphate	0.96	ns
Nitrate	0.62	0.72
Nitrite	0.55	0.75
Hardness	0.89	ns
Chloride	0.87	ns
CP	0.98	ns
HPC	0.82	0.50
TC	0.94	ns
Bacterial abundance	0.99	ns
Chlorophyll-*a*	0.96	ns
Explained variance (%)	64.86	20.36

2.3 Results

Sites selected in pristine location of the Morogoro Riverine system fulfilled twenty presumptive conditions of reference sites described by Sánchez-Montoya *et al*., (12) allowing the assessment of shifts in water quality as a result of influences from urban and agricultural land-uses. Except pH and dissolved oxygen which were more or less stable in the studied river, physico-chemical parameters showed fluctuations between land use types with significant shifts between pristine and urban or agricultural locations (Table 1, Fig. 1). However, the three land-use types could be significantly discriminated from each other by biological oxygen demand, ammonia, nitrate and

phosphate (all H-*test* p<0.05; Table 1). This observation was perfectly mirrored by chlorophyll-*a* concentration (Table 2) indicating the robustness of the chemical parameters in detecting extrinsic influences.

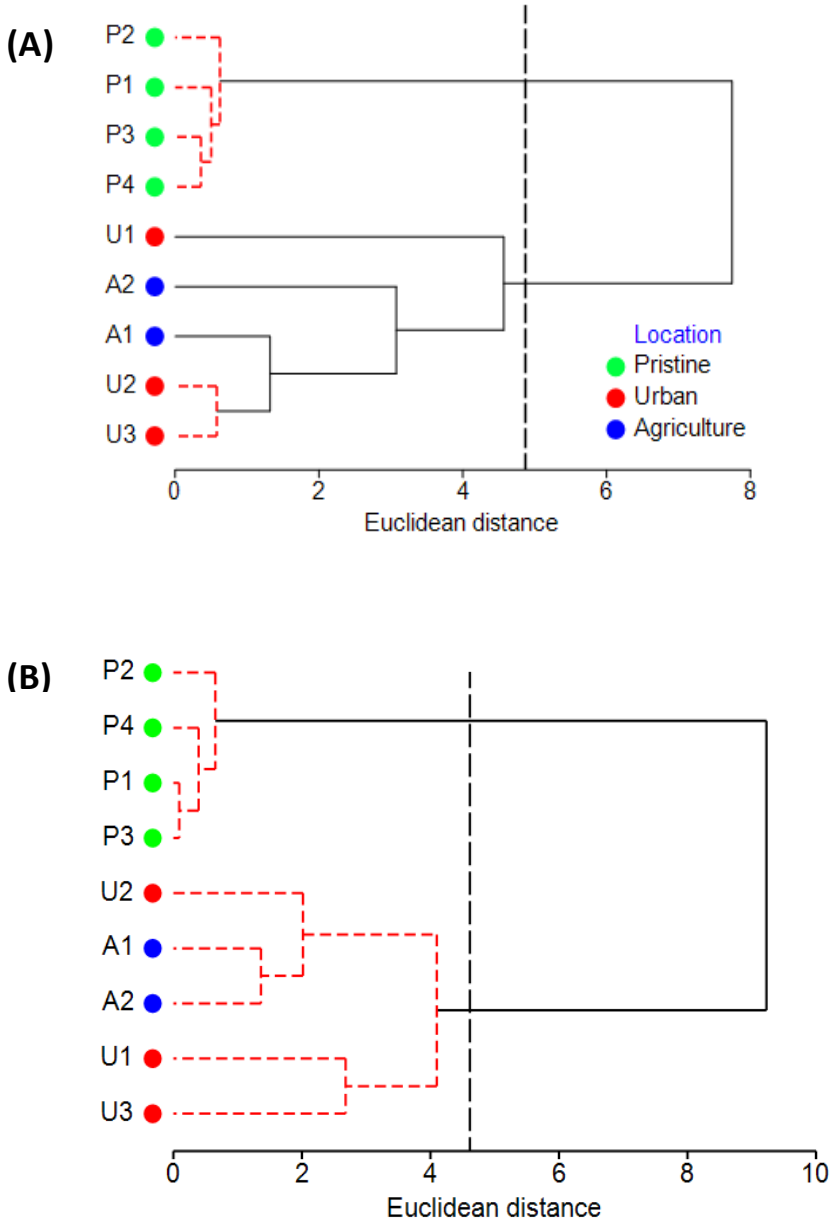


Figure 1: Classification of sampling points by cluster analysis using complete linkage, Euclidean distance and transformed (log+1) physico-chemical (**A**) and microbiological (**B**) data sets, respectively. Analysis was performed at 999 permutations. While red dashed lines within the cluster indicate those sites which were not separated by similarity profiles (SIMPROF) at *P*<0.05, the vertical black dashed line indicates significantly different clusters (*P*<0.05); n=90 single samples/measurements per parameter. P1 - P4 = pristine individual sites; U1 - U2 = urban individual sites, and A1 - A2 = agricultural individual sites.

Cluster analyses based on physico-chemical (Fig. 1A) and microbiological (Fig. 1B) Euclidean distances visibly showed that selected pristine sites were very similar to each other but very different from urban and agricultural sites indicating that watershed land-use had an effect on water quality variables. However, none of the cluster analyses could distinguish urban sites from agricultural sites because very few parameters varied between the two locations (Table 1 and 2) leading to in-ability of the applied multivariate statistical tests in detecting the differences. Evidence of distinct magnitude of disturbances were clearly featured by the dendogram generated from physico-chemical data set (Fig.1A) as Euclidean distances for the cluster containing urban and agricultural sites were very high and considerably variable as compared to that of pristine sites. However, the Euclidean distances generated from microbiological data for urban sites were not different from that of agricultural land use type (Fig 1B) indicating that microbiological variables responded differently to land use influences as compared to physico-chemical variables.

A verimax rotation of principal component analysis performed using the whole microbiological and physico-chemical data set revealed that riverine water quality exhibited perturbations that could be clearly explained by two major principal components whose eigenvalues were higher than 1 with a total of 85.2% explainable variations in the data set (Table 3). The first component (PC1) that explained 64. 86% of the variance suggested water quality deterioration caused by land use as variables which correlated strongly across, but not within, land use types formed part of this component. These included among others indicators of water quality deterioration (CP, TC, TSS, ammonia, BOD, nitrate) that are tightly linked to watershed land-use activities. In PC1, dissolved oxygen had a negative contribution whereas, pH had no contribution. Unlike PC1, the second component (PC2) accounted for 20.36% of the total variance with the highest contribution from pH, a variable which were not influenced by land use activities (Table 1). Presence of dissolved oxygen, conductivity, nitrate, nitrite and HPC in the principal component containing pH is evidence that at least part of these variables is contributed by naturally occurring processes in the riverine system. However, it should be noted that variation exhibited by these two components were not 100%, indicating that there exist other factor(s) apart from land-use and naturally occurring processes responsible for water quality deterioration, albeit to

a lesser extent, which could not be captured by variables detected in the studied riverine system.

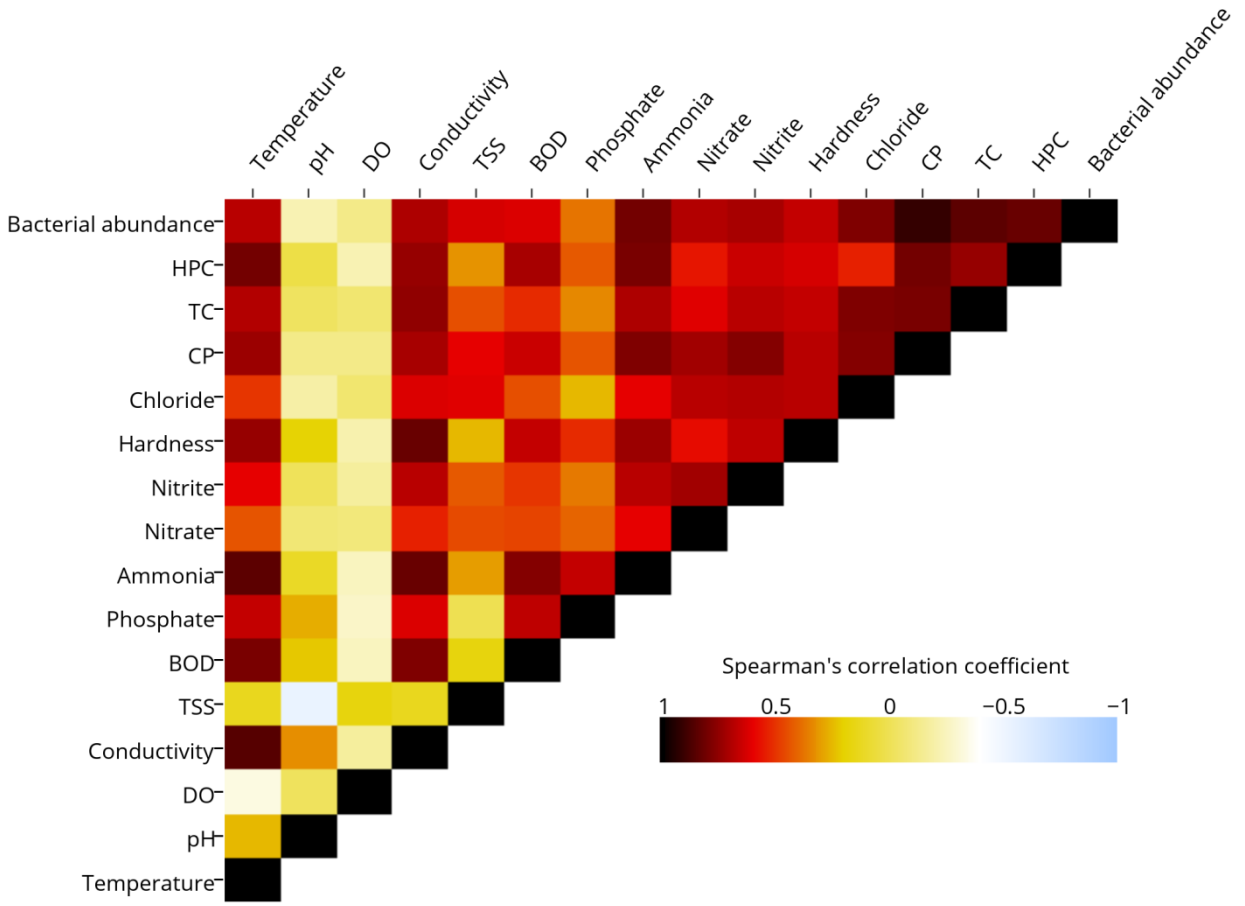


Figure 2: Correlation matrix of microbiological and physico-chemical variables determined in Morogoro Riverine system. Sampling was done 10 times in each site and all variables were separately determined from each sample resulting to a total of 10 values for each variable per site (n=90). For abbreviations see Table 1 and 2.

The Spearman correlation matrix showed that land-use gradients were clearly mirrored by both microbiological and physicochemical variables as high degree of correlation amongst parameters were only revealed when pooled data sets were considered. However, pH and dissolved oxygen were the only variables unaffected by land use (Table 1) and their coefficient of variations obtained following correlation with other variables were very low and non-significant (Fig 2). CP, allochtonous bacteria and an indicator of water deterioration, correlated strongly with other bacteriological parameters such as HPC, TC and bacterial abundance (Fig. 2)

and it was the only bacterial group which reflected the intermittent perturbations in the studied riverine system as revealed by the highest coefficient of variation within and across land use types (Table 2). CP could support the extrinsic origin of physico-chemical substances due to its substantial relationship with BOD, nitrate, phosphates and ammonia. Like Chlorophyll-*a*, CP statistically discriminated the three land use types from each other (Table 2).

2.4 Discussion

The present study identifies the impacts of watershed land-use on microbiological and physicochemical characteristics of tropical riverine system. Microbiological and physicochemical parameters displayed a distinct responsiveness to watershed land use (Table 1 & 2) and could prove that resilience was not possible during the sampling period suggesting that anthropogenic pollution overwhelmed self-purification (18).

No individual multivariate analysis provided a sufficient basis for land-use types classification (Fig. 1). Classification based on cluster analysis and microbiological or physico-chemical variables identified sites of low and high contaminations but did not discriminate between land-use types as many of the detected variables were not sensitive enough in responding to distinct land use influences. Consequently, multivariate method such as cluster analysis may not be appropriate for identification of urban, and agricultural land use influences. However, cluster analysis demonstrated that sampling sites in urban and agricultural land use types had a stable Euclidean distances (derived from microbiological data) whereas, the Euclidean distances derived from physico-chemical data for the same sites were substantially different from each other (Fig 1) indicating the importance of using both microbiological and physicochemical variables in riverine water quality assessment.

Data revealed that concentration of BOD, ammonia, phosphates and nitrates were tightly tied to the land-use type (Table 1). Although various studies have pointed out that bedrock of the river system can have a confounding effect on physicochemical variables (7,19,20), several lines of evidence suggested that anthropogenic influences detected in this study overwhelmed bedrock contributions. First, a clear correlation of these physicochemical parameters and CP

(originating from terrestrial soil- and gastro-intestinal tract of animals (21)) at an appreciable Spearman correlation coefficients ($r > 0.5$) was observed (Fig 2). Second, considerable differences of physicho-chemical parameters were detected between land-use types located in the same geological settings (see methodology section & Table 1). Lastly, 64.86% of the observed total variations (85.32%) in principal component analysis were explained by first component directly linked to anthropogenic influences (Table 3).

Of the biotic components tested, chlorophyll-*a* statistically detected differences between land-use types (Table 2). While its responsiveness to anthropogenic condition is obviously not unexpected given the detected levels of nutrients (nitrogen, phosphorus) in river water, its highest values coincide with urban influences including treated sewage effluents from oxidation ponds, whereas that from agricultural locations was a consequence of erosion due to poor livestock grazing practices. The fact that chlorophyll-*a* coincides strongly with the observed nutrient concentrations make this parameter a potential candidate for assessing the influences associated with watershed land use at large spatial scale.

CP considerably varied within and between land-use types highlighting the importance of stochastic extrinsic disturbances. There was a strong correlation of CP with bacterial abundance, HPC and TC (Fig 2) indicating that a substantial fraction of allochtonous bacteria dominated these bacterial groups. Although advection has been previously reported to increase bacterial dispersal by increasing the probability of bacterial cells from upstream to reach downstream sites of the river systems (22), many of the bacterial groups tested here did not show any difference between urban and agricultural sampling locations despite the former being located far upstream of the latter (Table 1). Similarly, the concentration of chlorophyll-*a* was higher in urban than in downstream agricultural sampling points, therefore previously reported advection could not interfere our observation.

Our study demonstrated that land-use was the main driver of riverine microbiological and physicochemical characteristics as proven by the outcome of principal component analysis (Table 3). This finding was supported by the fact that significant correlation of the detected variables were observed when pooled data of all land use types were considered, whereas

correlations of the same variables for individual land use data sets could lead to insignificantly very low Spearman coefficient values. Furthermore, contamination markers such as CP and TC remarkably formed a principal component (PC1) with variables that had increased values in urban and agricultural locations (Table 1; Table 3). As expected, dissolved oxygen had negative contribution in PC1 due to the high levels of organic matter contributed by urban and agricultural activities that consume large amount of oxygen as supported by high value of BOD. Besides land use, principal component two (PC2) showed that naturally occurring processes including microbial transformation of chemical substances may influence the physico-chemical quality of the investigated system, a finding similar to what have been reported elsewhere (23). The fact that we could not determine all parameters already proposed for water quality assessments, principal component analysis could not display 100% explainable variations. Therefore, it is possible that other factors driving the river water quality exist but with less effect as observed from the overall percentage of explainable variations by principal component analysis (Table 3).

The high contamination levels in various sites identified by sensitive physico-chemical and microbiological variables undermine livelihood, hinder sustainable development and diminish the ability of a watershed to perform its ecological work (2,18,24). Consequently, a watershed plan will be required that uses best-management practices and techniques that incorporate natural physical and biological processes to reduce, convert, or store pollutants on the land before they enter the aquatic system combined with a rational input management policy (25).

This study showed that land use diversity in the watershed and natural processes are the main determinant of microbiological and physico-chemical characteristics in a tropical riverine system. Individual physico-chemical variables including BOD, ammonia, phosphates, nitrates and microbiological variables such as Chlorophyll-*a* and CP were highly sensitive in statistically discriminating between dominant land use types observed in Morogoro River system and therefore may be promising candidates for effective water quality monitoring and evaluation of riverine system under restoration following land use influences.

2.5 Literature cited

1. Jiwa SF, Mugula JK, Msangi MJ. Bacteriological quality of potable water sources supplying Morogoro municipality and its outskirts: a case study in Tanzania. Epidemiol Infect. 1991;107:479–84.
2. Minaya V, McClain ME, Moog O, Omengo F, Singer GA. Scale-dependent effects of rural activities on benthic macroinvertebrates and physico-chemical characteristics in headwater streams of the Mara River, Kenya. Ecol Indic. 2013;32:116–22.
3. Ngoye E, Machiwa JF. The influence of land-use patterns in the Ruvu river watershed on water quality in the river system. Phys Chem Earth. 2004;29:1161–6.
4. Harding JS, Benfield EF, Bolstad P V, Helfman GS, Jones EB. Stream biodiversity: the ghost of land use past. Proc Natl Acad Sci USA. 1998;95:14843–7.
5. Verhoeven JTA, Arheimer B, Yin C, Hefting MM. Regional and global concerns over wetlands and water quality. Trends Ecol Evol. 2006;21(2):96–103.
6. Kazi TG, Arain MB, Jamali MK, Jalbani N, Afridi HI, Sarfraz RA., et al. Assessment of water quality of polluted lake using multivariate statistical techniques: A case study. Ecotoxicol Environ Saf. 2009;72(2):301–9.
7. Nelson ML, Rhoades CC, Dwire KA. Influence of bedrock geology on water chemistry of slope wetlands and headwater streams in the southern rocky mountains. Wetlands. 2011;31:251–61.
8. Singh KP, Malik A, Mohan D, Sinha S. Multivariate statistical techniques for the evaluation of spatial and temporal variations in water quality of Gomti River (India) - A case study. Water Res. 2004;38:3980–92.
9. Callender E, Hammond DE. Nutrient exchange across the sediment-water interface in the Potomac River estuary. Estuar Coast Shelf Sci. 1982;15(4):395–413.
10. Park JH, Duan L, Kim B, Mitchell MJ, Shibata H. Potential effects of climate change and variability on watershed biogeochemical processes and water quality in Northeast Asia. Environ Int. 2010;36:212–25.
11. Kirschner AKT, Kavka GG, Velimirov B, Mach RL, Sommer R, Farnleitner AH. Microbiological water quality along the Danube River: Integrating data from two whole-river surveys and a transnational monitoring network. Water Res. 2009;43:3673–84.
12. Sánchez-Montoya MDM, Arce MI, Vidal-Abarca MR, Suárez ML, Prat N, Gómez R. Establishing physico-chemical reference conditions in Mediterranean streams according to the European Water Framework Directive. Water Res. 2012;46:2257–69.
13. Pinto U, Maheshwari BL. River health assessment in peri-urban landscapes: An application of multivariate analysis to identify the key variables. Water Res. 2011;45:3915–24.
14. Gómez R, Arce IM, Sánchez JJ, del Mar Sánchez-Montoya M. The effects of drying on sediment nitrogen content in a Mediterranean intermittent stream: A microcosms study. Hydrobiologia. 2012;679:43–59.
15. Snelder TH, Biggs BJF. Multiscale river environment classification for water resources management. J Am Water Resour Assoc. 2002;38(5):1225–39.
16. Natkhin M, Dietrich O, Schäfer MP, Lischeid G. The effects of climate and changing land use on the discharge regime of a small catchment in Tanzania. Reg Environ Chang. 2013;DOI 10.1007/s10113-013-0462-2:1–12.

17. Gomani MC, Dietrich O, Lischeid G, Mahoo H, Mahay F, Mbilinyi B, et al. Establishment of a hydrological monitoring network in a tropical African catchment: An integrated participatory approach. Phys Chem Earth, Parts A/B/C. 2010;35(13-14):648–56.
18. Postel SL, Thompson BH. Watershed protection: Capturing the benefits of nature's water supply services. Nat Resour Forum. 2005;29:98–108.
19. Mosher JJ, Findlay RH. Direct and indirect influence of parental bedrock on streambed microbial community structure in forested streams. Appl Environ Microbiol. 2011;77(21):7681–8.
20. Holloway JM, Dahlgren RA, Hansen B, Casey WH. Contribution of bedrock nitrogen to high nitrate concentrations in stream water. Nature. 1998;395:785–8.
21. Shimizu T, Ohtani K, Hirakawa H, Ohshima K, Yamashita A, Shiba T, et al. Complete genome sequence of *Clostridium perfringens*, an anaerobic flesh-eater. Proc Natl Acad Sci USA. 2002;99(2):996–1001.
22. Ho DT, Schlosser P, Caplow T. Determination of longitudinal dispersion coefficient and net advection in the tidal Hudson River with a large-scale, high resolution SF6 tracer release experiment. Environ Sci Technol. 2002;36(15):3234–41.
23. Hartman WH, Richardson CJ, Vilgalys R, Bruland GL. Environmental and anthropogenic controls over bacterial communities in wetland soils. Proc Natl Acad Sci USA. 2008;105(46):17842–7.
24. Mathooko JM, M'Erimba CM, Kipkemboi J, Dobson M. Conservation of highland streams in Kenya: the importance of the socio-economic dimension in effective management of resources. Freshw Rev. 2009;2:153–65.
25. Osborne LL. Riparian vegetated buffer strips in water-quality restoration and stream management. Freshw Biol. 1993;29:243-258.

CHAPTER 3

Biogeography of tropical riverine bacterioplankton communities determined by deep sequencing of 16S rRNA gene amplicons

Mushi, D [1,2], Pereira, RPA[1], Höfle, MG[1] and Brettar I[1]

[1]Helmholtz Centre for Infection Research, Braunschweig (Germany)

[2]Sokoine University of Agriculture, P.O. Box 3038, Morogoro, Tanzania

3.0 Abstract

Riverine systems are intimately coupled with and shaped by the characteristics of their watershed, yet we know little about the effect of these connections to bacterioplankton communities. Here, we characterise water samples collected on the basis of dominant watershed land-use types observed along the tropical riverine system using 16S rRNA gene Illumina MiSeq sequencing. Three types of samples were analysed with different land-use types: pristine, urban and agricultural. Bacterioplankton communities were greatly discriminated by watershed land-use types rather than environmental or climate characteristics. Watershed land-use promoted allochtonous bacterial sequences, eutrophication-linked bacterial sequences, extinction and stress to indigenous bacterioplankton taxa. Despite the considerable variation across watershed land-use types, total bacterioplankton richness was greatly contributed by among-site richness than within-site richness. Bacterioplankton taxa showed a remarkable stability over time irrespective of climatic perturbations emphasizing the possibility of being temporally predictable.

3.1 Introduction

Bacterioplankton communities constitute a significant proportion of riverine biodiversity that facilitate many of the fundamental processes mediating ecosystem services, including biogeochemical nutrient cycling (1–5), energy flow (6,7), and pollutant removal (8). Of critical importance is a high sensitivity of bacterioplankton in responding to environmental alterations attributable to their ubiquity, small sizes, rapid growth rates and high abundance in the riverine system (9–13). Consideration of this potential phenomenon during riverine system monitoring provides useful information about the disturbance likely to have disrupted microbial activity, altered potential ecological processes, and influenced overall ecosystem performance (10,11,14–18). Recently, a number of studies have demonstrated distinct responses of bacterioplankton communities to changes happening in the river system (11,19,20). However, these responses were highly driven by intrinsic factors such as algal bloom and river impoundments, respectively. While there is a growing body of evidence that allochtonous

biogeochemical substances such as bacteria, nutrients and organic matter derived from watershed can enter the riverine system, we do not understand how they affect bacterioplankton communities given their importance in riverine ecology and within riverine biogeochemical processes (1–4,21). Consequently, a comprehensive analysis of interactions between extrinsic influences and bacterioplankton will allow the understanding of mechanisms driving river ecosystem which in turn enhance accurate prediction of disturbance, and sustainable management of this ecosystem.

Comprehensive and informative patterns of bacterioplankton assembly in extrinsically influenced riverine system are currently not yet available. A pronounced shift in riverine bacterioplankton assembly as a consequence of extrinsic influences has been recently recognized (16,22). However, these investigations focused exclusively on riverine biofilm and sediment related bacteria, respectively, despite the fact that bacterioplankton provides substantial information about recent processes happening in the water column of riverine system. Furthermore, a clear link between extrinsic influence and riverine bacterioplankton communities has been demonstrated (23,24). Unfortunately, these studies employed low-resolution molecular methods (DGGE and T-RFLP, respectively), low sampling efforts, and did not take into consideration the responses of individual taxa despite the fact that extrinsic influences could selectively favour specific taxa with potential information related to the ongoing ecological processes in the riverine water column.

We assessed the response of riverine bacterioplankton communities to extrinsic influences associated with distinct watershed land-uses by sequencing 16S rRNA gene amplicons at appropriate depth using the Illumina MiSeq platform. Water sampling sites were selected based on the observable dominant land-use types in the watershed of the respective tropical riverine system. These included sites selected in the tropical riverine system at a pristine location characterised by forest, those that were selected in the river at urban location, and those that were selected in the river at agricultural location, respectively. We hypothesize that bacterioplankton communities and diversity exhibit a pronounced spatial variation in tropical riverine system, variation that fundamentally reflects the watershed land-use patterns. We

explored whether the shift in relative abundance of specific taxa across land-uses or their extinction was associated with observed patterns of bacterioplankton communities. Taxa selectively favoured by extrinsic influences and their indication as a tool for possible riverine monitoring was examined. Also, we compared the contribution of among sites diversity and within site diversity to the total diversity based on inverse of Whittaker's beta diversity. To understand further the contribution of watershed land-use on riverine system, water column physicochemical characteristics were determined. We further hypothesize that bacterioplankton communities and physicochemical characteristics coincide strongly across watershed land-uses than within watershed land-use.

3.2 Materials and Methods

3.2.1 Study setting and investigational design

Morogoro riverine system belongs to the upper Ngerengere watershed geographically located between 6°51´, 7°09´S and 37°32´, 38°38´E and at altitude range of 500–2260m above sea level. The topography of the watershed ranged from the mountainous in the North-Western Uluguru mountains (2,260 m a. s. l.)–where the river originates–to the low land in the downstream reaches (25). The river extensively drains native forest of Uluguru Mountains, urban landscape at the foot of Uluguru Mountains, and livestock grazing land in the downstream reaches, respectively.

The tropical climate of the investigated watershed is characterized by ambient temperature, dropping below 20^{o}C in May and June. Annual rainfall ranges from 800 to more than 1500 mm with a bimodal regime (26). More than seventy percent of precipitation falls between March and May, the time when the hydrographs of the river indicate peak flows, whereas the remaining percentage accounted for the small peak that happens between November and April. Despite serving as a source of drinking water for the Cities of Morogoro and Dar es Salaam in Tanzania, the discharge in Morogoro riverine system ranged from 500 l/s in the headwaters to >3000 l/s in the downstream reaches during the time of this survey.

A total of 90 water samples was taken from 9 synoptic sampling points established along Morogoro River on the basis of adjoining watershed land-use types, and abundance of cultivable allochtonous bacteria (*Clostridium perfringens*, CP) inhabiting the water column. These included four synoptic sites established at pristine location, three sites at urban location, and two sites at agricultural location, respectively (see Fig. 2 in Chapter I). Watershed land use types were delineated according to Snelder and Biggs (27). Each sampling point was sampled 10 times during the complete study time that covered low and high flow regimes, and distinct degree of water column physico-chemical conditions (see Chapter 1).

The pristine landscape was located in the North-Western Uluguru Mountain within Uluguru nature reserve. While wild animals (such as vervet monkeys, wild pig, duiker antelopes, wild birds) and native forest dominated this location, neither human settlements nor anthropogenic activities were observed within and around this environment over the entire study period. The pristine headwaters were characterized by the lowest abundance of CP (median log 0.3 cfu per 100 ml) and flowed over Precambrian rocks of meta-sedimentary nature. Urban land-use adjoining Morogoro River is located immediately after the foot of Uluguru Mountain. A population of >300,000 people (Tanzania National Bureau of Statistics) inhabited this landscape. At this landscape, river water flowed over sediments of tertiary and quaternary ages. Domestic waste water and treated sewage from urban inhabitants and municipal sewage treatment plants, respectively are discharged to the river system. A median of log 1.3 cfu per 100 ml of CP was detected in the river water column at the urban location. Further downstream, agricultural land use dominated the watershed that involved livestock grazing (mainly cattle, goats and sheep) with direct access to river water especially during watering. Ingress of organic matter from the organic farming activities adjacent to the river water through erosion was also evident. Additionally, river water column at agricultural location was characterized with the highest CP approximating to the median of log 2.1 cfu per 100 ml. As in urban area, the river water at agricultural land-use flowed over the sediments of tertiary and quaternary ages.

3.2.2 Contextual meta data collection

Dissolved oxygen, temperature, conductivity and pH were assessed *in situ* with PCE-PHD 1 meter (PCE Deutschland, Meschede, Germany). Probes were calibrated at 25^{o}C prior to day of sampling and the calibration was substantiated after field measurements according to manufacturer´s instructions. Membrane electrode (4500-O G) and gravimetric techniques (2540D) were used to assay five day´s biochemical oxygen demand (BOD) and total suspended solids (TSS), respectively as described in American Public health Association standard methods (APHA 2000). Semi-quantitative QUANTOFIX® test strips (Macherey-Nagel, Düren, Germany) were used to estimate the concentration of ammonium (NH_4), phosphate (PO_4), nitrate (NO_3), nitrite (NO_2), chloride (Cl) ions and total hardness in the river water samples according to manufacturer's instructions.

3.2.3 Synoptic water sampling and pre-treatment

Water was sampled aseptically from the middle of the river, at a depth of approximately 30 cm, using 1 l wide-mouthed sterile plastic bottles (Thermo Scientific™ Nalgene™, Neubrecht, USA). Samples were immediately placed into a dark, ice-cooled (4°C) box and transported to Sokoine University of Agriculture (Morogoro, Tanzania) laboratory where bacterial cells from water samples were recovered and stored on FTA™ classic card (GE Healthcare UK Limited, Buckinghamshire, UK) within 6 to 9 hours after collection of the first sample. Briefly, about 250 to 2000 ml of water sample was filtered onto a 0.2-µm-pore-size polycarbonate filter (47 mm diameter; Nuclepore; Whatman, Maidstone, UK). After filtration, the filter was placed to a 47 mm sterile petri dish and kept wet with Milli-Q water. Bacterial cells were scraped off the filter surface as described elsewhere (28). Suspended cells in Milli-Q water were pipetted from the petri dish and transferred to the drawn circle area of the FTA™ classic card. The filter was rinsed twice with Milli-Q water, the rinsate was also transferred to the FTA™ classic card. Efficacy of retrieving bacterial cells from polycarbonate filters ranged from 99 to 100% (see suplementary Table S1 and associated information therein), suggesting that a desired number of cells were efficiently retrieved from the polycarbonate filters. FTA™ Classic card containing bacterial cell consortia were allowed to dry overnight on sterile silica gel before being placed into a labeled multi-barrier pouch (GE Healthcare Europe, Freiburg, Germany) to maintain

sample integrity during storage. Multi-barrier pouch containing samples were sealed and transported to Helmholtz Centre for Infection Research (HZI), Braunschweig, Germany for further processing. It should be highlighted that the FTA™ classic card was found to be the best tool for storing genomic DNA (gDNA) at room temperature for an extended period of time (> 2 months) without degradation following an experiment that compared several documented non-freezing methods for DNA preservation (see supplementary Fig. S1 and associated information therein).

3.2.4 Genomic DNA (gDNA) extraction

A total of 6 FTA punches (each with 6 mm diameter) containing DNA of the applied bacterial cells were punched out aseptically from FTA™ Classic cards using a metal single hole puncher and immediately placed into a sterile water bead tube (Mo Bio laboratories inc., Carlsbad, CA, USA). To avoid cross contamination of samples, the metal puncher was rinsed with 70% ethanol and flamed for 10s before being applied to a subsequent sample. gDNA was extracted from FTA punches using UltraClean™ PowerWater DNA isolation kit (Mo Bio laboratories inc., Carlsbad, CA, USA), following the manufacturer's instructions. The quantity of the extracted gDNA was gauged spectrophotometrically using NanoDrop 2000c spectrophotometer (NanoDrop Technologies, Wilmington, DE, USA) followed by quality analysis using 260/280 nm and 260/230 nm absorption ratios and agarose gel electrophoresis, respectively prior to storage at -20 ^{o}C for later use.

3.2.5 Library construction and deep sequencing of 16S rRNA gene amplicons

16S rRNA gene libraries were generated as described in Camarinha-Silva *et al*., (29) employing COM1F (5´-CAGCAGCCGCGGTAATAC-3´) and COM2R (5´-CCGTCAATTCCTTTGAGTTT-3´) universal primers that amplifies approximately 410 bp of V4 and V5 hypervariable regions of most bacterial 16S rRNA gene (30). Incorporated to the COM1F for each sample is the distinctive 6-base error correcting barcode and 2-base CA linker to allow for allocation of each sequence to the appropriate sample, and to avoid amplification bias (31,32), respectively. On the other hand, COM2R was fused with unique 6-base index to allow multiplexing of the samples. Both primer pairs contained appropriate adapters at the 5´ ends to permit sequencing on the

Illumina MiSeq platform (29,33). An amplification of 50 µl reaction mixture contained: 100 µM deoxynucleoside triphosphate (Bioline, Luckenwalde, Germany), 0.4 mM $MgCl_2$, 1× PCR reaction buffer, 0.03 U HotStarTaq Polymerase (Qiagen, Hilden, Germany), 0.4 µM of each primer and 2ng of gDNA template. PCR was performed on a Biorad Thermo cycler 96-well iCycler with an initial denaturation cycle of 95^{o}C for 15 min, followed by 30 cycles at 95^{o}C for 1 min, 55^{o}C for 40s, and 72 ^{o}C for 40s; a final extension of 10 min at 72^{o}C was added to ensure complete amplification. Two µl from the first PCR reaction mixture were used as template in a second PCR reaction performed under the same conditions as for the first PCR, except that 10 cycles, and PCR primers designed to integrate the sequence of the specific Illumina multiplexing sequencing primers and index primers were employed (29). Template free controls (using water instead of DNA) were performed in the two steps PCR procedure in parallel with the considered samples and, as they did not result in amplification (29), were not further considered in downstream amplicon processing.

To isolate 16S rRNA amplicons with approximately 410 base pair, PCR products were run on 2% agarose gel electrophoresis pre-stained with GelRed Nucleic Acid Gel Stain (Biotium, Hayward, CA, USA). Desired 16S rRNA bands on agarose gel were identified with the aid of blue light transilluminator. Bands were carefully excised and amplicons were extracted from the agarose gel according to the condition described in van Dijk *et al*., (34) using QIAquick Gel Extraction Kit (Qiagen, Hilden, Germany). The extracted amplicons were then quantified with Quant-iT PicoGreen dsDNA reagent (Life Technologies, Oregon, USA). Library tagged with unique index was prepared by pooling equimolar ratios of amplicons (≈30 ng of each sample) from an appropriate number of samples. Prior to deep sequencing, normalized libraries were purified using QIAGEN MinElute® PCR purification Kit (Qiagen, Hilden, Germany), quality determined, and quantified using Agilent BioAnalyzer platform. Sequencing of the quantified libraries was performed at genome analytics facility (GMAK) of the Helmholtz Centre for Infection Research (HZI), Braunschweig, Germany using the Illumina® MiSeq platform.

3.2.6 16S rRNA gene sequence data analysis

16S rRNA sequences generated by Illumina MiSeq platform were processed using Mothur software package version 1.27.0 (35). Briefly, the 4 028 396 raw sequences obtained from 36 gDNA samples sequenced by Illumina MiSeq platform were trimmed to remove primer and barcode sequences, and to eliminate poor quality sequences such as sequences shorter than 200 bp, those that contained ambiguous bases greater than 2%, those that comprised 2% nucleotide homopolymers, and those that had a quality score less than 40 in a sliding window of 50 nucleotides over the sequence length. Residual clean sequences were screened for chimeras using UCHIME algorithm (36) in Mothur and sequences suspected to be chimeras were excluded from downstream analysis. High quality sequences (3 042 461) that persisted a train of stringent quality control were aligned against the SILVA data base Incremental Aligner (37). Sequences were assigned to different operational taxonomic units (OTUs) at 97% sequence similarity by using Mothur and the RDP taxonomic data base release 9 (38).

3.2.7 Statistical assays

Analysis of similarity (ANOSIM; (39) based on Bray-Curtis similarity coefficient (40) was used to test the extent of the difference exhibited by bacterioplankton communities between sampling locations using PRIMER version 7.0.6 software (PRIMER-E, Plymouth, UK). Non-parametric Kruskal Wallis test was used to explore the differences in relative abundance of the detected bacterioplankton taxa among the locations, and the taxa responsible for the observed differences were identified by White's non-parametric test with adjusted probability value according to Bonferroni correction criterion. Plots of pairwise statistical comparisons according to Whites's test (41) were performed in the STAMP software package (42). Bacterioplankton community structure was performed based on Bray-Curtis similarity coefficient (40) and visualized by canonical analysis of principal coordinates plot (CAP, (43). Alpha diversity metrics were performed by using the Explicet software (44) after standardization of bacterial operational taxonomic units (OTUs) to 33,200 sequences to account for differences in the number of sequences of the considered libraries. The relationship between individual taxa and physicochemical variables were explored using Spearman correlation algorithm whereas, the relationship between microbial communities with either individual or whole physicochemical

data set was performed by using Mantel-like RELATE test that takes on board the Spearman correlation algorithm in the PRIMER version 7.0.6 software.

3.3 Results

3.3.1 Riverine bacterioplankton community diversity

Between October 2012 and June 2013, a total of 36 gDNA samples covering a range of environmental characteristics such as high and low flow regimes, dominant watershed land-uses, and distinct degree of physico-chemical conditions was collected from the water column of the Morogoro River in Tanzania. PCR amplification of 16S rRNA gene portion covering V4 and V5 hypervariable regions and subsequent deep sequencing of amplicons at an appropriate sequencing depth using Illumina MiSeq platform generated a total of 3, 042, 461 high quality 16S rRNA gene sequences from 36 gDNA samples with an average length of 410 nucleotides. Sequences ranging from 33 423 to 142 298 with an average of 84 513 per sample were detected in individual samples. Although amplicons were normalized prior to deep sequencing, the number of sequences were not the same across samples (see Fig. 1), most likely due to distinct bacterioplankton composition among samples. By clustering the high quality sequences to operational taxonomic units (OTUs) at 97% sequence identity threshold led to a total of 1,980 non-singleton OTUs. Good's coverage averaging to 99.5% ± 0.0013% (Mean±SD; range = 93.75-99.94%) for individual samples and over 99.5% per land use was observed. A collector's curve based on non-singleton OTUs per gDNA sample suggested that sampling efforts exceeding the analysed 36 gDNA samples will not lead to additional bacterioplankton OTUs (supplementary Fig. S2). Non-singleton OTUs per individual sample ranged from 283 to 1,063 (average of 692 ± 239 per sample) whereas, per land use the range was from 1,429 to 1,614 OTUs. Sequencing efforts applied to the broad range of analysed samples proved to be appropriate in resolving all the OTUs needed for complete representation of investigated riverine bacterioplankton diversity as demonstrated by asymptotic collector's curves generated from sequence abundances and observed OTUs data set at the scale of individual site (n=4), land-use (n=8-16), and whole catchment landscape (n=36) (Fig. 1).

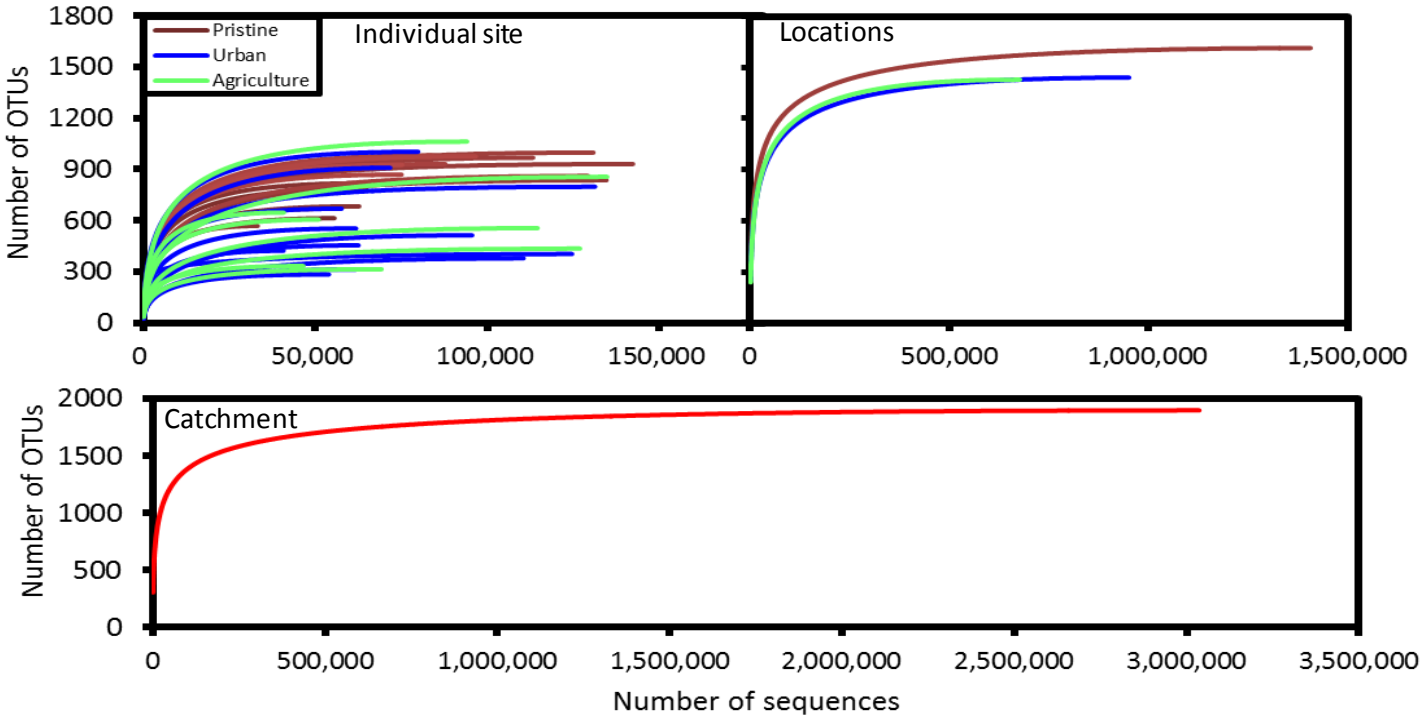


Figure 1: Collector's curve of observed bacterial OTU richness for the individual sampling site, locations and catchment scales. 16S rRNA OTUs were clustered conservatively at 97% sequence similarity.

To explore the dynamics of bacterioplankton richness along the investigated riverine system, OTUs were rarefied to the same number of sequences (33,200 sequences/site) to standardize sequencing efforts. The rarefied OTUs were substantially different amongst land-uses (Kruskal Wallis test, $P<0.001$) with higher number of OTUs in pristine as compared to that of urban and agricultural locations (see supplementary Fig S3). On closer examination, the inverse of Whittaker's beta diversity whose values ranged from 0.42 to 0.52 suggested that total OTUs detected in each independent location were highly contributed by amongst sites richness than within site richness.

3.3.2 Phylogenetic composition of riverine bacterioplankton communities

The high quality 16S rRNA gene sequences (3, 042, 461 sequences) from 36 gDNA samples were highly diverse and could be classified in more than 30 phyla affiliated to bacterial (sequences, 99.73%) and archaeal (0.27%) domains. In spite of this diverse microbiome, very few phyla

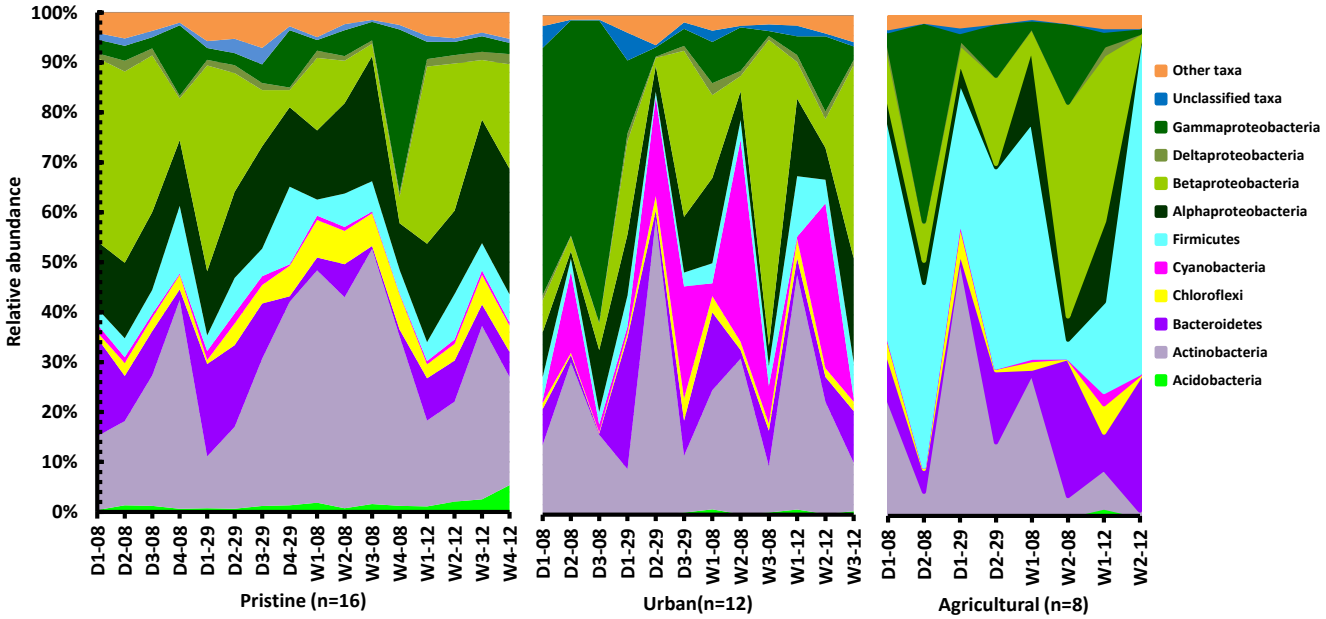


Figure 2: Relative abundance of dominant phyla across thirty six (36) gDNA samples collected from pristine (n=16), urban (n=12) and agricultural (n=8) locations in River Morogoro, Tanzania. Unidentified sequences were aggregated as 'unclassified taxa'. Phyla with relative abundance < 6% across all samples were aggregated into a single group named 'other taxa'. The codes for each water sample are presented along the X-axis and indicate the water sampling time of the year (D=Dry; W=Wet), sample identification number (1, 2, 3, 4) per season, and sampling dates (8.10.2012, 29.10.2012, 8.05.2013, 12.06.2013), y-axis represent the percentage of sequences for each dominant phylum detected in the considered gDNA samples.

tended to be dominant (accounting for 89.61 to 99.22% of the total sequences) and prevalent in all samples, albeit in substantially varied abundances (Fig. 2), demonstrating the presence of indigenous core bacteria in tropical riverine microbiome. Relative abundance ranging from 1% to 59% in individual samples were affiliated to the phylum *Actinobacteria*, 0.08 - 8% to *Chloroflexi*, 0.24 - 28% to *Bacteroidetes*, 0.02 - 41% to *Cyanobacteria*, 1 - 66% to *Firmicutes*, and 3 - 79% to *Proteobacteria*. Unlike cosmopolitan phyla, *Acidobacteria* was consistently detected in all gDNA samples taken from pristine location, whereas it was only occasionally present in urban and agricultural samples (Fig. 2). The remaining phyla (>24 phyla) and unclassified sequences were occasionally detected in the analysed samples regardless of sample origin, and together made up 1.54% of the total sequences detected in this study. Sequences of phylum to genus level were independently correlated between technical triplicates (mean Pearson correlation coefficient = 0.9924±0.013; range = 0.9453-0.9994; $P<0.001$; n = 9 independent sets

of triplicates). Figure S4 shows the sequence abundance of identified phyla in the independent sets of technical triplicates assessed in this investigation.

Bacterioplankton taxa showed varying relative abundance that mirrored the presumed response to dominant watershed land-uses around the investigated riverine system (Fig. 3 and Fig S5). Across all three land-use types, *Chloroflexi* and *Alphaproteobacteria* showed higher relative abundance in pristine than in urban and agricultural locations. The *Cyanobacteria* abundance was higher in urban land-use than in pristine and agricultural locations. Unlike samples taken from pristine and urban locations, agricultural samples had the highest relative abundance for *Firmicutes* (Fig. 3).

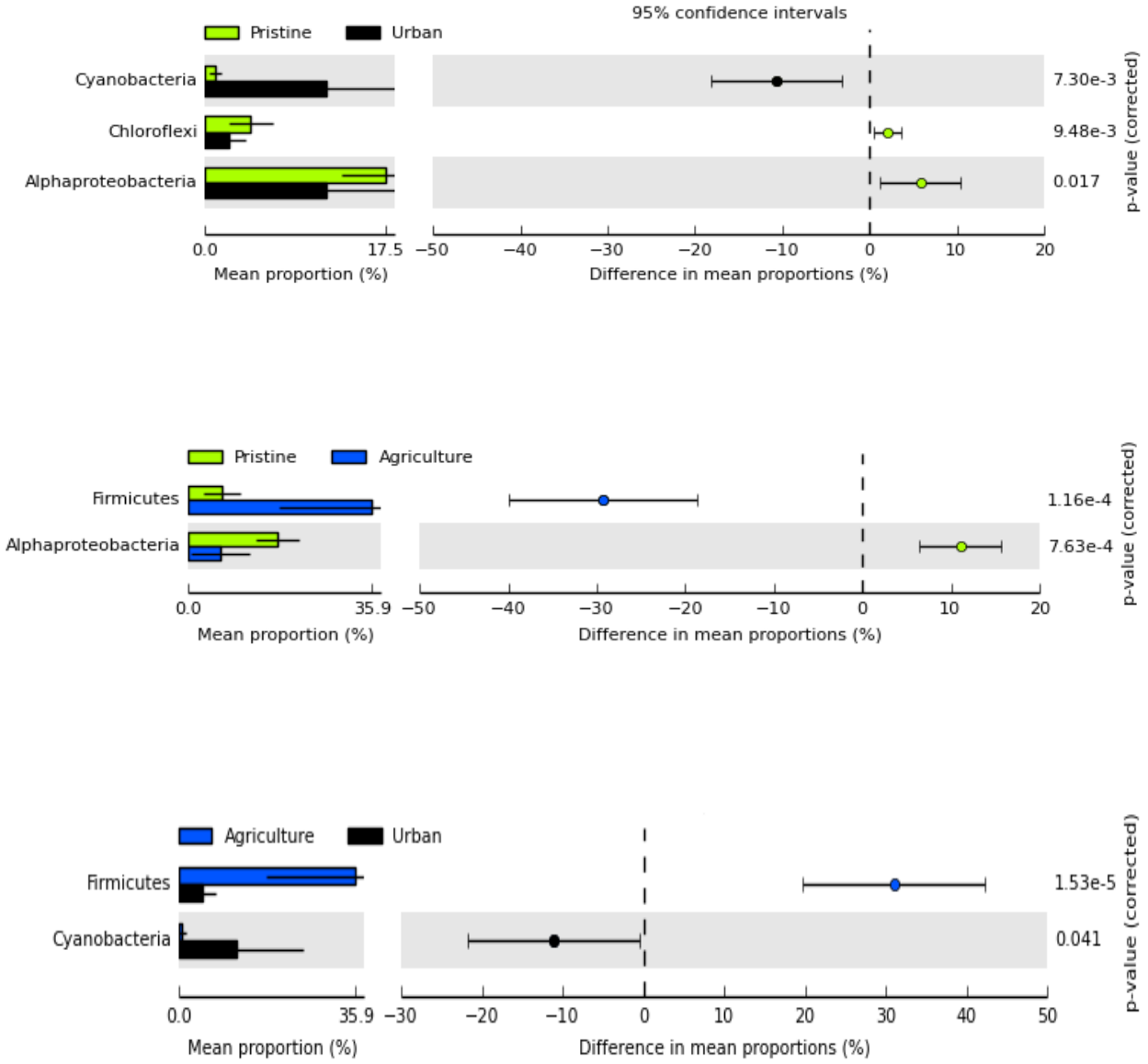


Figure 3: Pairwise comparison of mean sequence abundance (%) at phylum level between locations based on White's statistical test (White *et al*. 2009) and 95% confidence interval (CI). Note that only those phyla which showed statistical significant difference between the compared pair of locations are displayed here.

The evidence that the pristine river water inoculates bacterial taxa into other locations downstream of the investigated riverine system such as that dominated by urban and agricultural land uses was demonstrated by the presence of a substantial high number of core bacterial taxa across the 36 analysed samples (supplementary Fig. S6). However, a small fraction of bacterial OTUs disappeared, presumably in correspondence to water quality changes along the river system. OTUs detected frequently in pristine water were not found in the rest of the locations despite the adequate sampling and sequencing efforts employed in this study. These OTUs were affiliated to the genera *Acrocarpaspora*, *Chthonomonas*, *Fibrella, Porticoccus*, *Diplorickettsia*, *Agreia*, *Compostimonas*, and *Virgisporangium* among others. On the other hand, unique OTUs were detected in urban and agricultural locations that were never present upstream in the pristine location. OTUs affiliated to the genus *Stella*, Pir1 lineages, and SAR116 clade were detected only at urban sites, whereas those belonging to *Denitrobacterium*, *Caryophanon*, *Murdochiella*, *Sporanaerobacter*, *Paenalcaligenes* and *Thiobaca* were exclusively present at agricultural sites.

To further understand the connection between watershed and riverine system, specific taxa exclusively representing a particular habitat in the watershed were examined from the detected riverine bacterial assemblages. From this assessment, OTUs affiliated to genus *Chthonomonas*, *Clostridium*, *Bacillus*, *Arthrobacter, Anaeromyxobacteria*, *Azotobacter*, *Azospirillum* and *Actinomedura* that are exclusively found in soil habitat (45–54) were detected in the water column of the investigated riverine system. On the other hand, the river water column contained OTUs affiliated to the genera *Victivallis, Blautia*, *Bifidobacteria*, *Enterococcus and Akkermasia* that are members of animal intestinal microbiome (55–60). OTUs affiliated to the genera *Gaiella and Geothrix* commonly found in groundwater aquifer (61,62) formed part of the riverine bacterioplankton consortia, whereas OTUs affiliated to *Agreia*, *Clavibacter*, *Frankia, Rhizobium* and *Bradyrhizobium* (63–67) that are known to form symbiosis with roots of terrestrial plant were also detected in this tropical river water.

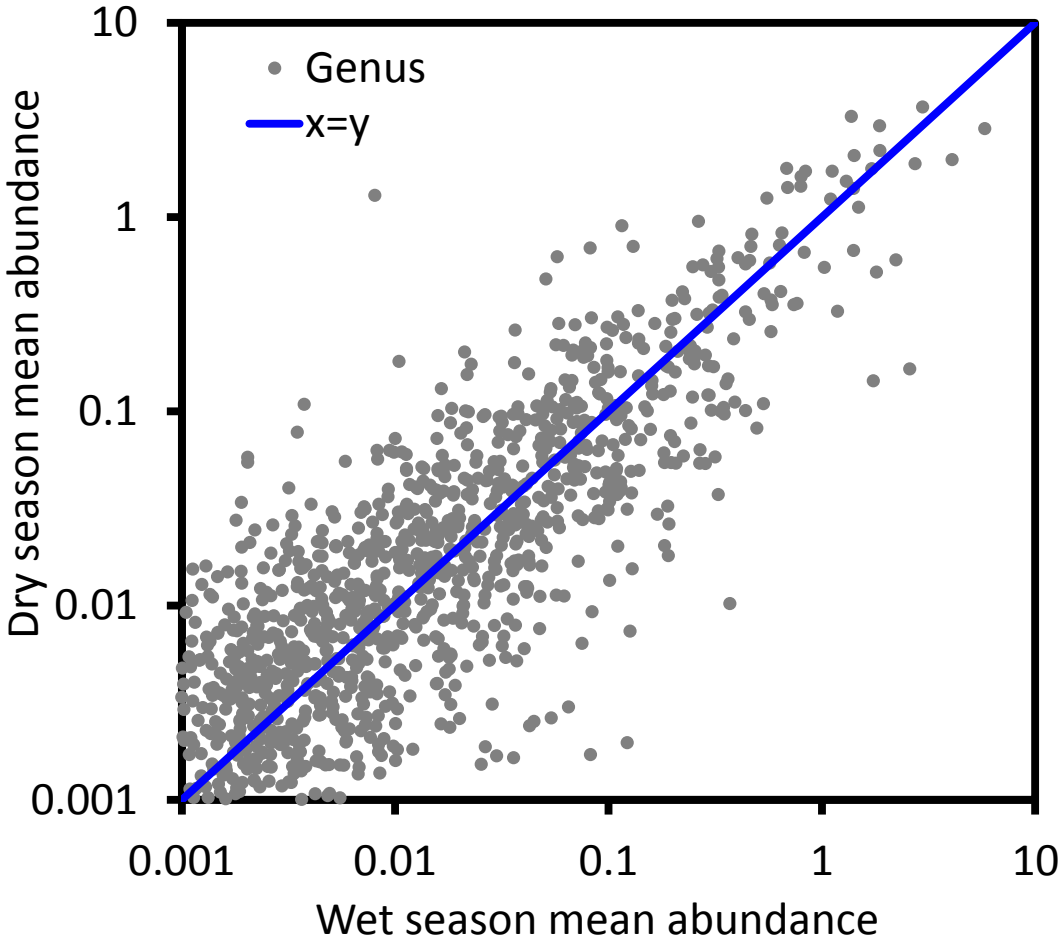


Figure 4. Correlation between mean relative abundance (log-scaled) data set of the wet season (x-axis) and that of the dry season (y-axis)–determined after an interval of six months–for the detected bacterial genera with mean relative abundance >0.001. It should be noted that genera instead of OTUs were used in this analysis for clarity. Each point in the plot indicates one genus and the position of the genus in the plot reflect the relative abundance obtained during wet (x-value) and dry (y-value) seasons. While spearman rank correlation coefficient between the two sampling occasions were highly significant (rho = 0.86; $p<0.001$), the difference between seasons were not significant ($P>0.05$). The blue diagonal shows the 1:1 ratio of the relative abundances of the respective pair of seasons.

3.3.3 Presence and abundance of bacterial taxa in tropical riverine system is temporally predictable

The sampling compaign performed between October and November, 2012, and between May and June, 2013 unveiled bacterioplankton (at genus level) data sets that were mirror image of each other (Fig. 4) despite an interval of six months between the two sampling occasions that were characterised by different climatic conditions (dry and wet seasons, respectively). Over

90% of the bacterioplankton genera had relative abundances that could not be statistically distinguished (*t*-test, P>0.05) between sampling occasions, and correlation of relative abundance of genera between these distinct sampling occasions were statistically significant (Spearman rho = 0.86; p<0.001). However, there were very few outliers that could show a significant difference (P<0.05) between sampling occasions probably the most sensitive genera whose relative abundance are shaped by rainfall events (See Fig. 4).

3.3.3 Bacterioplankton community variability

Although seasonal fluctuations in hydrology were evident during the course of riverine system survey, their impact on riverine microbiome was not detectable because bacterioplankton community profiles of dry (low water level) and wet (high water level) events were noticeably very similar (Global ANOSIM *R* = 0.02; *P*>0.05). In contrary, bacterioplankton community profiles were grouped by their watershed landuse with distinct pristine, urban and agricultural bacterioplankton assemblages (Fig. 5) as shown by Bray Curtis-based canonical analysis of principal coordinate (43). These bacterioplankton assemblages were significantly distinct from one another (Global ANOSIM *R* = 0.60; *P* = 0.001, 999 permutations) as proven by non-parametric one way analysis of similarity (ANOSIM; supplementary Table S2). Pairwise comparison revealed that pristine bacterioplankton assemblage was distinct from that found in urban (ANOSIM *R* = 0.5, *P* = 0.001), and agricultural land-uses (ANOSIM *R* = 0.83; *P*=0.001). Similarly, urban and agricultural bacterioplankton assemblages were quite distinct (ANOSIM *R* = 0.51; *P* = 0.001). Noticeably, ANOSIM values generated from Sørensen-based presence-absence algorithm (Global *R* = 0.38; *P*=0.001; 999 permutations) using the same bacterioplankton community profile data set were as significant as that performed by using abundance-based Bray-Curtis (Global ANOSIM *R* = 0.60; *P* = 0.001, 999 permutations) which demonstrated that taxonomic patterns observed here (Fig 5) were substantially influenced by differences in relative abundance of the detected bacterial taxa across the land-uses as well as taxa extinction. This statistical outcome complements the observed distinctions in relative abundance for those phyla (see Fig. 2 and 3) and bacterial order (supplementary Fig. S5) that tended to strongly respond to a particular land-use but also the observed substantial decline of number of OTUs across watershed land-use (supplementary Fig. S3).

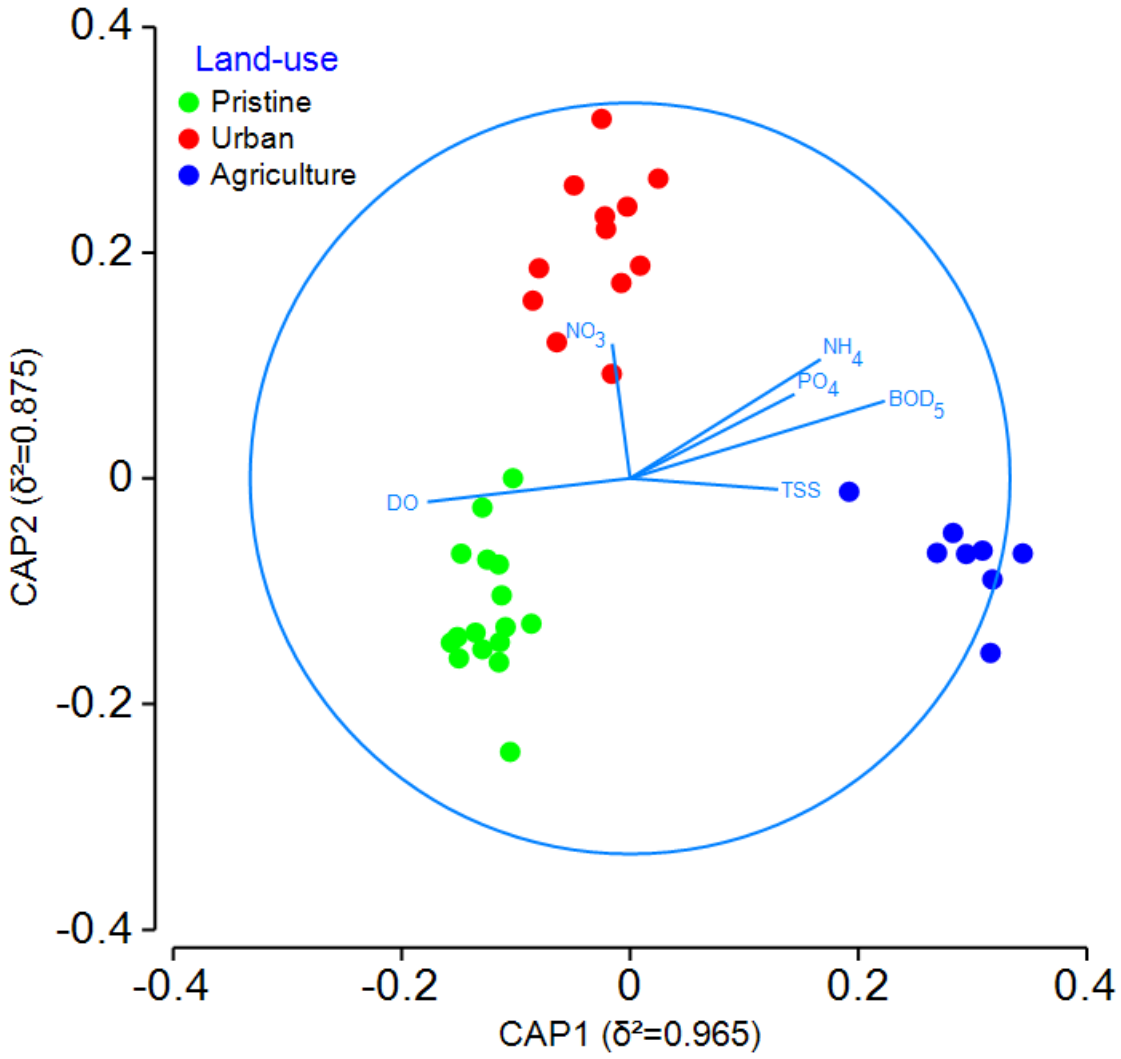


Figure 5. Canonical analysis of principal coordinates (CAP) plot illustrating the axes that best discriminate bacterioplankton communities according to watershed land-uses in the surveyed river system. Taxa abundance (% of sequences) data set was standardized according to Clarke *et al*., (68) before applying Bray-Curtis similarity coefficient (40) for the generation of the ordination plot. Each bacterioplankton community profile was assigned to the cluster for which it had the highest probability of membership (69). Bacterioplankton community profiles are shown as dot symbols: green dots represent pristine-associated bacterioplankton assemblages, red dots represent urban-associated bacterioplankton assemblages, and blue dots represent agriculture-associated bacterioplakton assemblages. Vectors of selected environmental variables were overlaid on CAP plot after determination of bacterioplankton community structure, and are proportional to their correlation with either CAP1 axis or CAP2 axis. The circle of correlation is displayed in which those environmental variables with vectors closer to the circle of correlation had higher correlation coefficient than those closer to the center. The canonical correlation (δ^2) of each CAP axis indicates the strength of multivariate data and hypothesis of differences between locations (69). The CAP reclassification rates for pristine, urban and agriculture were 93.7%, 75% and 87.5%, respectively indicating the extent of discrimination among the locations realized by the canonical axes (69). The sum of canonical eigenvalues (traceQ_m´HQ_m statistic) was 1.84 at *P*=0.001 (999 permutations). Abbreviations: DO, dissolved oxygen; BOD, five day´s biological oxygen demand; TSS, total suspended solids.

3.3.4 Relationship between bacterioplankton communities and contextual meta data

Direct interaction between bacterioplankton communities and a broad spectrum of water column physico-chemical characteristics were performed using Mantel-like RELATE test. Correlation between global physico-chemical data set (composite of 12 variables) and bacterioplankton communities (composite of 36 bacterioplankton community profiles) was evident (RELATE $R = 0.61$, $p < 0.001$). However, bacterioplankton communities displayed a varying relationship with individual physico-chemical variables in which biological oxygen demand and ammonium had the highest correlation coefficients ($R = 0.62$ and 0.69, respectively; $p < 0.001$) followed by temperature ($R=0.57$, $p=0.001$), conductivity ($R=0.49$, $p=0.001$), total hardness ($R= 0.44$, $p = 0.001$), and phosphates ($R = 0.49$, $p = 0.001$). Although significant, nitrite ($R = 0.39$; $p = 0.001$), dissolved oxygen ($R=0.29$; $p=0.003$), total suspended solids ($R = 0.26$; $p = 0.004$), chlorides ($R= 0.18$; $p = 0.003$) and nitrates ($R= 0.16$; $p = 0.018$) had weak correlations, whereas pH did not correlate with bacterioplankton communities (supplementary Table S3). The considered range of differently influenced samples spanning pristine, urban and agricultural land use types supports a strong correlation observed between bacterioplankton communities and various environmental variables. Interestingly, correlation between bacterioplankton communities and individual environmental variables within each location was statistically negligible ($R<0.1$; $p>0.05$).

On the other hand, correlation of each individual taxa (at order level) and environmental variables showed interesting patterns (supplementary Fig. S7). *Actinobacteria*_PeM15 showed significant positive correlation with pH, whereas *Clostridiales* were the only taxa that showed significant positive correlation with total suspended solids, biological oxygen demand and ammonium. The hypothesis that relative abundances of several taxa were significantly altered across the land-use (see Fig. 3, 5 and S5) was clearly featured in the correlation analyses, as much as 13 bacterial orders showed significant negative correlation with at least one of the environmental variable that had increased values from pristine through urban to agricultural land-use (supplementary Figure S7).

3.4 Discussion

To completely comprehend the tropical riverine bacterioplankton diversity and its pattern of spatial distribution, appropriate sampling– and 16S rRNA gene amplicon sequencing efforts are of paramount importance. Here, we showed that analysis of appropriate number of gDNA samples followed by sequencing of 16S rRNA gene amplicons at adequate depth generated operational taxonomic units (OTUs) representing the complete bacterioplankton diversity of the investigated tropical river water (Fig. 1 and S2). These OTUs (>1,980 OTUs) were affiliated with >30 known phyla (Fig. 4), indicating that the tropical riverine system constituted as high bacterioplankton diversity as that reported in riverine systems occurring in other geographical regions (11,70). The majority of the diverse sequences (89-99%) detected were affiliated with *Proteobacteria*, *Firmicutes*, *Cyanobacteria*, *Chloroflex*, *Bacteroidetes* and *Actinobacteria* that were also detected in previous studies on riverine systems (7,70–72), demonstrating that riverine systems constitute ubiquitous core bacterioplankton regardless of their geographical location, molecular technique employed, and DNA extraction method applied. Despite the high number of gDNA samples analyzed that spanned distinct seasons, watershed land-use types, and physicochemical conditions, these predominant bacterioplankton phyla tended to be cosmopolitan (dominant phyla were detected across all samples including those taken from headwaters and downstream reaches in this study) representing the robust nature of the tropical riverine system in maintaining key bacterioplankton phyla despite the frequent disturbance, and emphasize the importance of headwaters as the primary source of these phyla.

Our investigation revealed that watershed land-use patterns significantly shaped individual bacterioplankton abundance (Fig. 3). Agricultural land-use substantially enhanced proportions of *Firmicutes* in the investigated riverine system, an increase deemed of direct input because members of this bacterial group can be considered as allochtonous in river water (73,74). Unlike *Firmicutes*, substantial increase of *Cyanobacteria* abundance in riverine system was correlated with treated sewage

effluents from urban land-use whose contributions to *Cyanobacteria* development are well established (75–77). However, the substantial reduction in relative abundance of *Chloroflex* and *Alphaproteobacteria* sequences in the urban and agricultural locations of the studied river system (Fig. 3) was somewhat surprising considering the phototrophic nature of the former (14) and involvement of the latter in carbon cycling (78). However, this inter-location variation in relative abundance of *Chloroflex* and *Alphaproteobacteria* is evidence that occupancy of different niches in the riverine system with different linkages to other taxa results to different responses to disturbance (79). Similar observation was experienced elsewhere (14) and intolerable harsh condition (stress) was suggested as a possible causes of this variation. These responses provide the first evidence that watershed land-use have a range of effects on riverine ecosystem that can be accurately realized by individual bacterioplankton taxa. Consequently, these bacterioplankton taxa could be used as sensors to detect the presence of disturbances related to direct bacterial input, eutrophication and ecological stress that may evolve from watershed land-use such as that observed in the investigated tropical river water. Since no technical variation was detected during this survey (see the results section and Fig. S4), the bacterioplankton variations observed here can be confidently considered to have caused by watershed land-use influences.

Significant change in relative abundance of specific bacterioplankton across watershed land-uses (Fig. 2 and 3), extinction of certain bacterioplankton taxa in urban and agricultural locations (Fig. S3), and presence of bacteria originating from terrestrial environment (see the results section) provide clear evidence that the riverine system has a strong interaction with watershed land-use. This interaction was further supported by the remarkable patterns of bacterioplankton community structure that was mirror image of watershed land-use patterns observed in the investigated tropical riverine systems (Fig. 5). This tight connection indicates the robust nature of bacterioplankton communities in detecting the extrinsic influences emanating from watershed land-uses and provided a clear evidence that the documented bias

associated with multi-template PCR reactions (80) did not affect the presumed dynamics of bacterioplankton communities in the investigated riverine system. It should be noted that the extent of discrimination among the locations realized by the canonical axes were not 100% instead between 75% to 93.7% (see caption of Fig. 5) meaning that apart from watershed land-use there exist other factors that influenced the structure of bacterioplankton in the tropical river system but to a lesser extent.

Despite the fact that sample collection spanned major seasonal events (dry and wet) known to strike the tropical environment, robust spatial patterns of bacterioplankton communities driven by watershed land-uses extremely obscured the influence of these seasonal events. While this observation is in broad agreement with findings in other riverine system elsewhere (21), the sampling design employed in this study may have favored more the spatial patterns than seasonal influences given that the land-use types considered here are not closely tied to seasons and that even the abiotic factors affecting bacterioplankton composition were highly altered by land-use rather than seasons. Importantly, the bacterioplankton data set revealed that the influence of seasonality in the investigated riverine system can only be captured by a limited number of certain individual taxa of bacterioplankton (see Fig. 4) rather than community structure or bacterioplankton diversity. Remarkably, bacterioplankton of the studied tropical river system were qualitatively and quantitatively reproducible between sampling occasions (Fig. 4) of six months interval irrespective of rainfall disturbance leading to the hypothesis that bacterioplankton of tropical riverine systems are predictable over time.

Several studies have documented numerous local factors potentially driving bacterioplankton community structure in freshwater systems, such as water chemistry, pH, temperature, and salinity (1,13,81–83). However, in this study such environmental factors did not correlated with bacterioplankton communities within each independent sampling location ($R<0.1$; $p>0.05$) despite the high sampling efforts per location, instead the significant correlations were observed across locations (Table S3) with most of the

correlation coefficients being much lower than 0.5 (except BOD and ammonia) indicating that most of these factors have relatively little contribution to the observed bacterioplankton community structure and emphasize that correlation across locations were highly driven by land-use influences. This finding was highly supported by the fact that bacterioplankton communities had a high association with ammonium and biological oxygen demand that are known to be signatures of anthropogenic influences (84–87). Correlations at the lower taxonomic level (bacterial order) could clearly verify that land-use was the main driver of bacterioplankton in tropical river water as allochtonous *Clostridiales* were the only bacterial order which had significant positive correlation with environmental variables emanating from watershed land-use (84,85,87) i.e. ammonia, biological oxygen demand and total suspended solids.

In summary, multiple lines of evidence suggested watershed land-use as the main driver of bacterioplankton communities in tropical riverine system with as high as 75% to 93.7% discrimination power which was not observed with single or combined environmental variables or seasonal influences evaluated here. Watershed land-use had substantial influence on presence- and relative abundance of bacterioplankton taxa. Interaction between riverine and terrestrial ecosystems, eutrophication signature, and presence of biological stress were revealed by specific bacterioplankton taxa that are recommended for use in monitoring the riverine system as far as watershed land-use is concerned. Environmental variables reported previously as drivers of bacterioplankton communities were substantially shaped by watershed land-use and showed very low variation with bacterioplankton. Bacterioplankton richness showed substantial difference between watershed land-use types, and that among-site diversity contributed more to total diversity than within-site diversity. Tropical riverine bacterioplankton showed a remarkable stability over several months in spite of climatic perturbations and emphasized that these riverine bacterioplankton communities are temporally predictable.

3.5 Literature cited

1. Hahn MW. The microbial diversity of inland waters. Curr Opin Biotechnol. 2006;17(3):256–61.
2. Azam F, Worden AZ. Microbes, molecules, and marine ecosystems. Science. 2004;303:1622–24.
3. Cotner JB, Biddanda BA. Small players, large role: Microbial influence on biogeochemical processes in pelagic aquatic ecosystems. Ecosystems. 2002;5:105–21.
4. Battin TJ, Kaplan LA, Findlay S, Hopkinson CS, Marti E, Packman AI, et al. Biophysical controls on organic carbon fluxes in fluvial networks. Nat Geosci. 2009;1:95–100.
5. Read DS, Gweon HS, Bowes MJ, Newbold LK, Field D, Bailey MJ, et al. Catchment-scale biogeography of riverine bacterioplankton. ISME J. 2015;9:516–26.
6. Sinsabaugh RL, Findlay S, Franchini P, Fischer D. Enzymatic analysis of riverine bacterioplankton production. Limnol Oceanogr. 1997;42(1):29–38.
7. Crump BC, Peterson BJ, Raymond PA, Amon RMW, Rinehart A, McClelland JW, et al. Circumpolar synchrony in big river bacterioplankton. Proc Natl Acad Sci USA. 2009;106(50):21208–12.
8. Sweeney BW, Bott TL, Jackson JK, Kaplan LA, Newbold JD, Standley LJ, et al. Riparian deforestation, stream narrowing, and loss of stream ecosystem services. Proc Natl Acad Sci USA. 2004;101(39):14132–7.
9. Zeglin LH. Stream microbial diversity in response to environmental changes: review and synthesis of existing research. Front Microbiol. 2015;6:454.
10. Solé M, Fetzer I, Wennrich R, Sridhar KR, Harms H, Krauss G. Aquatic hyphomycete communities as potential bioindicators for assessing anthropogenic stress. Sci Total Environ. 2008;389:557–65.
11. Yan Q, Bi Y, Deng Y, He Z, Wu L, Van Nostrand JD, et al. Impacts of the Three Gorges Dam on microbial structure and potential function. Sci Rep. 2015;5:8605.
12. Sun MY, Dafforn KA, Brown MV, Johnston EL. Bacterial communities are sensitive indicators of contaminant stress. Mar Pollut Bull. 2012;64:1029–38.
13. Logue JB, Lindström ES. Biogeography of bacterioplankton in inland waters. Freshw Rev. 2008;1:99–114.
14. Drury B, Rosi-Marshall E, Kelly JJ. Wastewater treatment effluent reduces the abundance and diversity of benthic bacterial communities in urban and suburban rivers. Appl Environ Microbiol. 2013;79(6):1897–905.
15. Staley C, Gould TJ, Wang P, Phillips J, Cotner JB, Sadowsky MJ. Bacterial community structure is indicative of specific types of contamination in the Upper Mississippi River. Appl Environ Microbiol. 2013;115(5):1147–58.
16. Lear G, Lewis GD. Impact of catchment land use on bacterial communities within stream biofilms. Ecol Indic. 2009;9(5):848–55.
17. Loza V, Perona E, Mateo P. Molecular fingerprinting of cyanobacteria from river biofilms as a water quality monitoring tool. Appl Environ Microbiol. 2013;79(5):1459–72.
18. Wakelin SA, Colloff MJ, Kookana RS. Effect of wastewater treatment plant effluent on microbial function and community structure in the sediment of a freshwater stream with variable seasonal flow. Appl Environ Microbiol. 2008;74(9):2659–68.

19. Winter C, Hein T, Kavka G, Mach RL, Farnleitner AH. Longitudinal changes in the bacterial community composition of the Danube River: A whole-river approach. Appl Environ Microbiol. 2007;73(2):421–31.
20. Ruiz-González C, Proia L, Ferrera I, Gasol JM, Sabater S. Effects of large river dam regulation on bacterioplankton community structure. FEMS Microbiol Ecol. 2013;84:316–31.
21. Portillo MC, Anderson SP, Fierer N. Temporal variability in the diversity and composition of stream bacterioplankton communities. Environ Microbiol. 2012;14(9):2417–28.
22. Gibbons SM, Jones E, Bearquiver A, Blackwolf F, Roundstone W, Scott N, et al. Human and environmental impacts on river sediment microbial communities. PLoS One. 2014;9(5):e97435
23. Levine UY, Crump BC. Bacterioplankton community composition in flowing waters of the Ipswich River watershed. Biol Bull. 2013;203:251–2.
24. Mark Ibekwe A, Leddy MB, Bold RM, Graves AK. Bacterial community composition in low-flowing river water with different sources of pollutants. FEMS Microbiol Ecol. 2012;79:155–66.
25. Natkhin M, Dietrich O, Schäfer MP, Lischeid G. The effects of climate and changing land use on the discharge regime of a small catchment in Tanzania. Reg Environ Chang. 2013;DOI 10.1007/s10113-0462-2.
26. Gomani MC, Dietrich O, Lischeid G, Mahoo H, Mahay F, Mbilinyi B, et al. Establishment of a hydrological monitoring network in a tropical African catchment: An integrated participatory approach. Phys Chem Earth Parts A/B/C. 2010;35(13-14):648–56.
27. Snelder TH, Biggs BJF. Multiscale river environment classification for water resources management. J Am Water Resour Assoc. 2002;38(5):1225–39.
28. Barbeau K, Kujawinski EB, Moffett JW. Remineralization and recycling of iron, thorium and organic carbon by heterotrophic marine protists in culture. Aquat Microb Ecol. 2001;24:69–81.
29. Camarinha-Silva A, Jáuregui R, Chaves-Moreno D, Oxley APA, Schaumburg F, Becker K, et al. Comparing the anterior nare bacterial community of two discrete human populations using Illumina amplicon sequencing. Environ Microbiol. 2014;16(9):2939–52.
30. Schwieger F, Tebbe CC. A New approach to utilize PCR−single-strand-comformation polymorphism for 16S rRNA gene-based microbial community analysis Appl Environ Microbiol. 1998;64(12):4870–6.
31. Hamady M, Walker JJ, Harris JK, Gold NJ, Knight R. Error-correcting barcoded primers for pyrosequencing hundreds of samples in multiplex. Nat Methods. 2008;5(3):235–7.
32. Kittelmann S, Seedorf H, Walters WA, Clemente JC, Knight R, Gordon JI, et al. Simultaneous amplicon sequencing to explore co-occurrence patterns of bacterial, archaeal and eukaryotic microorganisms in rumen microbial communities. PLoS One. 2013;8(2).e47879
33. Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Huntley J, Fierer N, et al. Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. ISME J. 2012;6:1621–4.
34. Van Dijk EL, Jaszczyszyn Y, Thermes C. Library preparation methods for next-generation sequencing: Tone down the bias. Exp Cell Res. 2014;322(1):12–20.

35. Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, et al. Introducing mothur: Open-source, platform-independent, community-supported software for describing and comparing microbial communities. Appl Environ Microbiol. 2009;75(23):7537–41.
36. Edgar RC, Haas BJ, Clemente JC, Quince C, Knight R. UCHIME improves sensitivity and speed of chimera detection. Bioinformatics. 2011;27(16):2194–200.
37. Pruesse E, Peplies J, Glöckner FO. SINA: Accurate high-throughput multiple sequence alignment of ribosomal RNA genes. Bioinformatics. 2012;28(14):1823–9.
38. Cole JR, Wang Q, Cardenas E, Fish J, Chai B, Farris RJ, et al. The Ribosomal Database Project: Improved alignments and new tools for rRNA analysis. Nucleic Acids Res. 2009;37:141–5.
39. Clarke KR. Non-parametric multivariate analyses of changes in community structure. Aust J Ecol. 1993;18:117–43.
40. Bray JR, Curtis JT. An Ordination of the upland forest community of southern Wisconsin. Ecology Monographs. 1957;27(4):325–49.
41. White JR, Nagarajan N, Pop M. Statistical methods for detecting differentially abundant features in clinical metagenomic samples. PLoS Comput Biol. 2009;5(4).e1000352
42. Parks DH, Beiko RG. Identifying biologically relevant differences between metagenomic communities. Bioinformatics. 2010;26(6):715–21.
43. Anderson MJ, Wallis TJ. Canonical Analysis of Principal Coordinates : a useful method of constrained ordination for ecology. 2003;84(2):511–25.
44. Robertson CE, Harris JK, Wagner BD, Granger D, Browne K, Tatem B, et al. Explicet: Graphical user interface software for metadata-driven management, analysis and visualization of microbiome data. Bioinformatics. 2013;29(23):3100–1.
45. Lee KCY, Dunfield PF, Morgan XC, Crowe MA, Houghton KM, Vyssotski M, et al. Chthonomonas calidirosea gen. nov., sp. nov., an aerobic, pigmented, thermophilic micro-organism of a novel bacterial class, Chthonomonadetes classis nov., of the newly described phylum Armatimonadetes originally designated candidate division OP10. Int J Syst Evol Microbiol. 2011;61:2482–90.
46. Janssen PH. Identifying the Dominant Soil Bacterial Taxa in Libraries of 16S rRNA and 16S rRNA Genes. Appl Environ Microbiol. 2006;72(3):1719–28.
47. Busse H, Wieser M. the genus Arthrobacter. IN: Rosenberg E, Delong EF, Lory S, Stackebrandt E, Thompon F: The Prokaryotes.Berlin, Springer, pp. 105-132. ISBN:978-3-30137-7.
48. Miranda-Ríos JA, Ramírez-Trujillo JA, Nova-Franco B, Beltran LFL, Iturriaga G, Suárez-rodríguez R. Draft genome sequence of *Arthrobacter chlorophenolicus* strain Mor30.16, Isolated from the bean rhizosphere. 2015;3(3):e00360-15.
49. Sanford RA, Cole JR, Tiedje JM. Characterization and description of *Anaeromyxobacter dehalogenans* gen. nov., Sp. nov., an aryl-halorespiring facultative anaerobic myxobacterium. Appl Environ Microbiol. 2002;68(2):893–900.
50. Thomas SH, Wagner RD, Arakaki AK, Skolnick J, Kirby JR, Shimkets LJ, et al. The mosaic genome of *Anaeromyxobacter dehalogenans* strain 2CP-C suggests an aerobic common ancestor to the delta-proteobacteria. PLoS One. 2008;3(5).e2103
51. Slepecky RA., Hemphill HE. The Genus Bacillus—Nonmedical. Prokaryotes Vol. 4. 2006. 530-562 p.

52. Tejera N, Lluch C, Martínez-Toledo M V, González-López J. Isolation and characterization of *Azotobacter* and *Azospirillum* strains from the sugarcane rhizosphere. Plant Soil. 2005;270:223–32.
53. Bashan Y. Inoculants of plant growth-promoting bacteria for use in agriculture. Biotechnol Adv. 1998;16(4):729–70.
54. Zakharova OS, Zenova GM, Zvyagintsev DG. Some approaches to the selective isolation of actinomycetes of the genus *Actinomadura* from soil. Microbiology. 2003;72(1):126–30.
55. Zoetendal EG, Plugge CM, Akkermans ADL, de Vos WM. Victivallis vadensis gen. nov., sp. nov., a sugar-fermenting anaerobe from human faeces. Int J Syst Evol Microbiol. 2003;53:211–5.
56. Derrien M, Vaughan EE, Plugge CM, de Vos WM. Akkermansia municiphila gen. nov., sp. nov., a human intestinal mucin-degrading bacterium. Int J Syst Evol Microbiol. 2004;54(5):1469–76.
57. Derrien M, Collado MC, Ben-Amor K, Salminen S, De Vos WM. The mucin degrader Akkermansia muciniphila is an abundant resident of the human intestinal tract. Appl Environ Microbiol. 2008;74(5):1646–8.
58. Eren AM, Sogin ML, Morrison HG, Vineis JH, Fisher JC, Newton RJ, et al. A single genus in the gut microbiome reflects host preference and specificity. ISME J. 2015;9:90-100.
59. Bottacini F, Medini D, Pavesi A, Turroni F, Foroni E, Riley D, et al. Comparative genomics of the genus Bifidobacterium. Microbiology. 2010;156:3243–54.
60. Byappanahalli MN, Nevers MB, Korajkic A, Staley ZR, Harwood VJ. Enterococci in the environment. Microbiol Mol Biol Rev. 2012;76(4):685–706.
61. Albuquerque L, França L, Rainey FA, Schumann P, Nobre MF, Da Costa MS. *Gaiella occulta* gen. nov., sp. nov., a novel representative of a deep branching phylogenetic lineage within the class *Actinobacteria* and proposal of *Gaiellaceae* fam. nov. and *Gaiellales* ord. nov. Syst Appl Microbiol. 2011;34(8):595–9.
62. Coates JD, Ellis DJ, Gaw C V, Lovley DR. Geothrix ferrnentans gen. nov., sp. nov., a novel Fe(III)-reducing bacterium from a hydrocarborn-contaminated aquifer. Int J Syst Bacteriol. 1999;49:1615–22.
63. Evtushenko LI, Dorofeeva LV, Dobrovolskaya TG, Streshinskaya GM, Subbotin SA, Tiedje JM. *Agreia bicolorata* gen. nov., sp. nov., to accommodate actinobacteria isolated from narrow reed grass infected by the nematode *Heteroanguina graminophila*. Int J Syst Evol Microbiol. 2001;51:2073–9.
64. Jacques MA, Durand K, Orgeur G, Balidas S, Fricot C, Bonneau S, et al. Phylogenetic analysis and polyphasic characterization of *Clavibacter michiganensis* strains isolated from tomato seeds reveal that nonpathogenic strains are distinct from *C. michiganensis* subsp. michiganensis. Appl Environ Microbiol. 2012;78(23):8388–402.
65. Normand P, Orso S, Cournoyer B, Jeannin P, Chapelon C, Dawson J, et al. Molecular phylogeny of the genus *Frankia* and related genera and emendation of the family Frankiaceae. Int J Syst Bacteriol. 1996;46(1):1–9.
66. Zahran HH. *Rhizobium*-legume symbiosis and nitrogen fixation under severe conditions and in an arid climate. Microbiol Mol Biol Rev. 1999;63(4):968–89.
67. Rivas R, Martens M, de Lajudie P, Willems A. Multilocus sequence analysis of the genus Bradyrhizobium. Syst Appl Microbiol. 2009;32(2):101–10.

68. Clarke KR, Chapman MG, Somerfield PJ, Needham HR. Dispersion-based weighting of species counts in assemblage analyses. Mar Ecol Prog Ser. 2006;320:11–27.
69. Hartmann M, Frey B, Mayer J, Mäder P, Widmer F. Distinct soil microbial diversity under long-term organic and conventional farming. ISME J. 2015;9:1177–94.
70. Jackson CR, Millar JJ, Payne JT, Ochs CA. Free-Living and Particle-Associated Bacterioplankton in Large Rivers of the Mississippi River Basin Demonstrate Biogeographic Patterns. Appl Environ Microbiol. 2014;80(23):7186–95.
71. García-Armisen T, Inceoğlu Ö, Ouattara NK, Anzil A, Verbanck MA, Brion N, et al. Seasonal variations and resilience of bacterial communities in a sewage polluted urban river. PLoS One. 2014;9(3):e92579.
72. Staley C, Unno T, Gould TJ, Jarvis B, Phillips J, Cotner JB, et al. Application of Illumina next-generation sequencing to characterize the bacterial community of the Upper Mississippi River. J Appl Microbiol. 2013;115:1147–58.
73. Hu A, Yang X, Chen N, Hou L, Ma Y, Yu CP. Response of bacterial communities to environmental changes in a mesoscale subtropical watershed, Southeast China. Sci Total Environ. 2014;472:7460–756.
74. Marti E, Balcázar JL. Use of pyrosequencing to explore the benthic bacterial community structure in a river impacted by wastewater treatment plant discharges. Res Microbiol. 2014;165(6):468–71.
75. Jarvie HP, Neal C, Withers PJA. Sewage-effluent phosphorus: A greater risk to river eutrophication than agricultural phosphorus? Sci Total Environ. 2006;360(1-3):246–53.
76. Van Drecht G, Bouwman a. F, Harrison J, Knoop JM. Global nitrogen and phosphate in urban wastewater for the period 1970 to 2050. Global Biogeochem Cycles. 2009;23(GB0A03):1–19.
77. Loza V, Perona E, Mateo P. Specific responses to nitrogen and phosphorus enrichment in cyanobacteria: Factors influencing changes in species dominance along eutrophic gradients. Water Res. 2014;48(1):622–31.
78. Fierer N, Bradford MA, Jackson RB. Toward an ecological classification of soil bacteria. Ecology. 2007;88(6):1354–64.
79. Dethlefsen L, Relman DA. Incomplete recovery and individualized responses of the human distal gut microbiota to repeated antibiotic perturbation. Proc Natl Acad Sci USA. 2011;108 (Suppl1):4554–61.
80. Pinto AJ, Raskin L. PCR biases distort bacterial and archaeal community structure in pyrosequencing datasets. PLoS One. 2012;7(8).e43093
81. Pearce DA, van der Gast CJ, Woodward K, Newsham KK. Significant changes in the bacterioplankton community structure of a maritime Antartic freshwater lake following nutrient enrichment. Microbiology. 2005;151:3237–48.
82. Lindström ES, Agterveld MPK, Lindstro ES, Zwart G. Distribution of Typical Freshwater Bacterial Groups Is Associated with pH , Temperature , and Lake Water Retention Time Distribution of Typical Freshwater Bacterial Groups Is Associated with pH , Temperature , and Lake Water Retention Time. Appl Environ Microbiol. 2005;71(12):8201–6.
83. Tang X, Xie G, Shao K, Bayartu S, Chen Y, Gao G. Influence of salinity on the bacterial community composition in Lake Bosten, a large oligosaline lake in arid Northwestern China. Appl Environ Microbiol. 2012;78(13):4748–51.

84. Radwan M, Willems P, El-Sadek A, Berlamont J. Modelling of dissolved oxygen and biochemical oxygen demand in river water using a detailed and a simplified model. Int J River Basin Manag. 2003;1(2):97–103.
85. Coulter CB, Kolka RK, Thompson JA. Water quality in agricultural, urban, and mixed land use watersheds. J Am Water Resour Assoc. 2004;40(6):1593–601.
86. Baker A. Land use and water quality. Hydrological process. 2003;17:2499-501.
87. Berka C, Schreier H, Hall K. Linking water quality with agricultural intensification in a rural watershed. Water Air Soil Pollut. 2001;127:389–401.

3.6 Supplementary materials

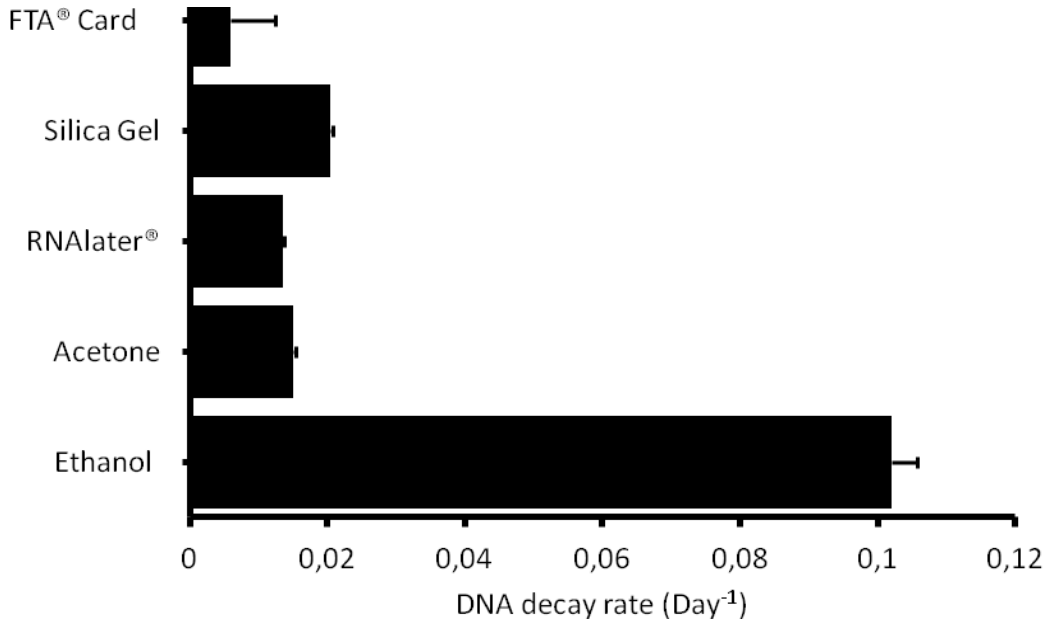


Figure S1: Comparisons of non-freezing DNA storage methods with respect to DNA decay over time (>8 weeks) at room temperature. Compared methods were previously applied in various studies including Rissanen *et al.*, (2010) (Ethanol and RNAlater®), Fukatsu (1999) (Acetone), Bainard *et al.*, (2010) (Silica gel) and Saieg *et al.* (2012) (FTA Cards). Of all the methods compared, DNA on FTA card had the lowest decay rate averaging to 0.006 per day.

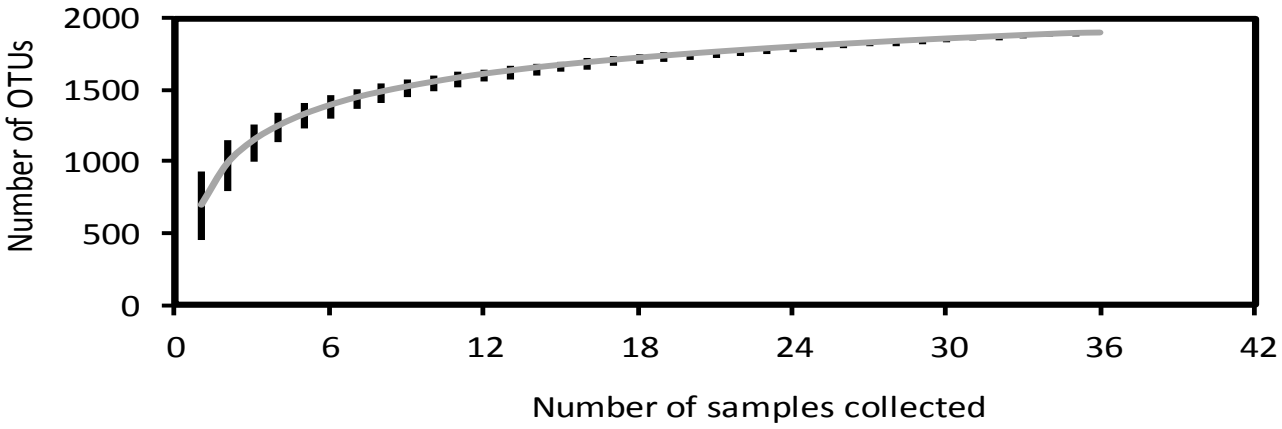


Figure S2: Collector's curve (±SD) of observed bacterial OTU richness against 36 gDNA samples considered in this survey. 16S rRNA OTUs clustered conservatively at 96% sequence similarity.

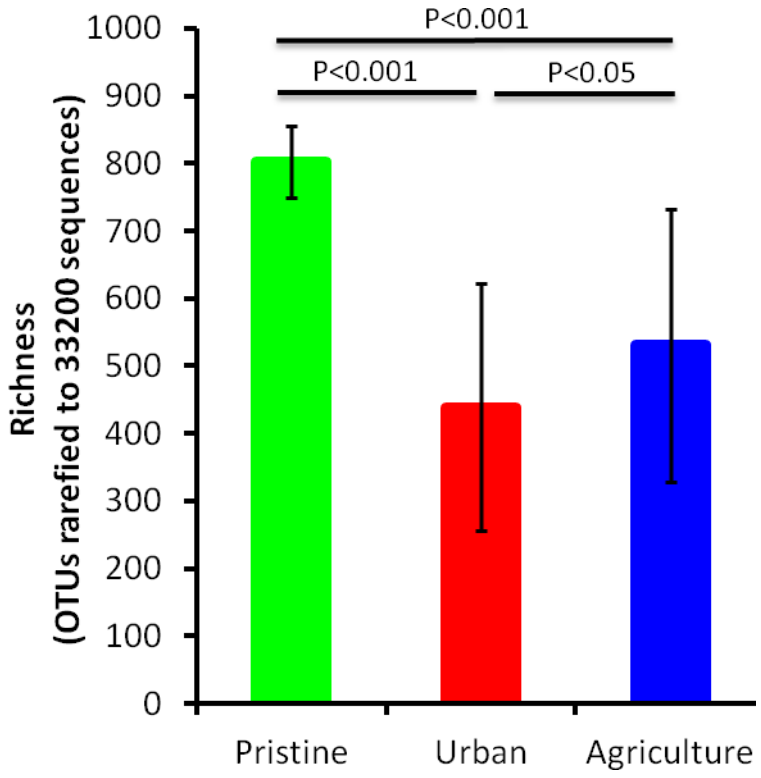


Figure S3: Comparison of bacterioplankton richness between watershed land-use of the surveyed tropical river system. Of the analysed samples (n=36 gDNA samples), pristine, urban and Agriculture had 16, 12, 8 number of gDNA samples, respectively.

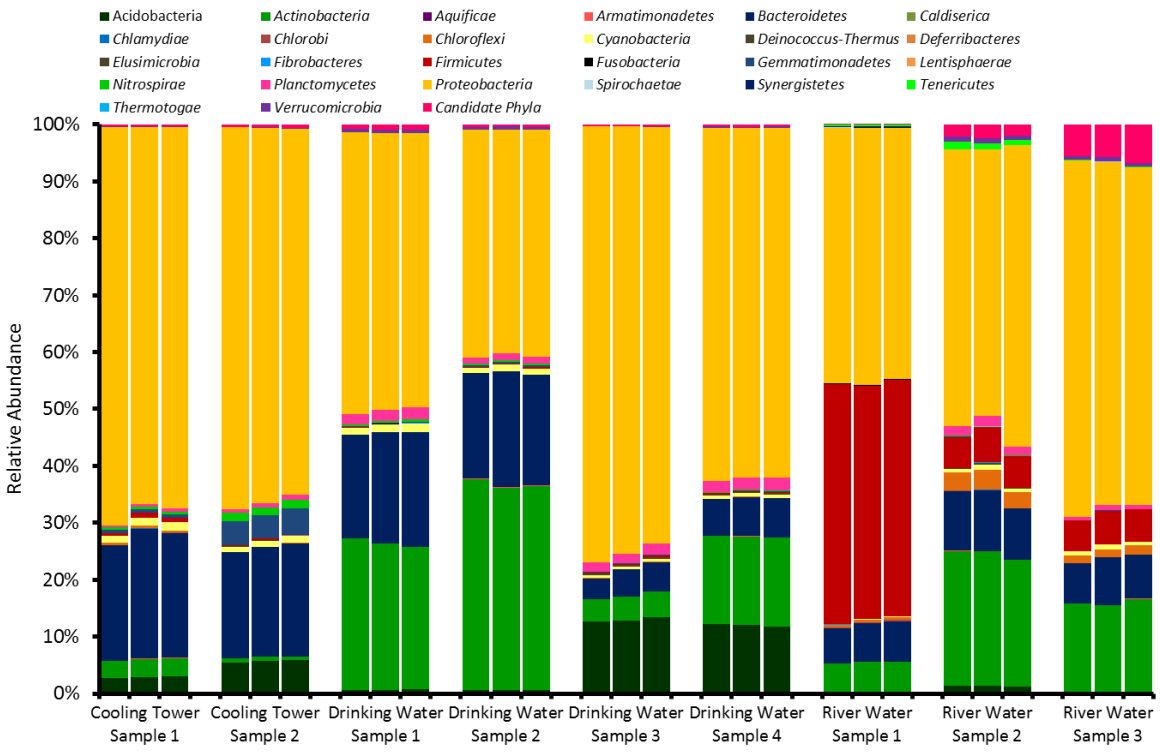
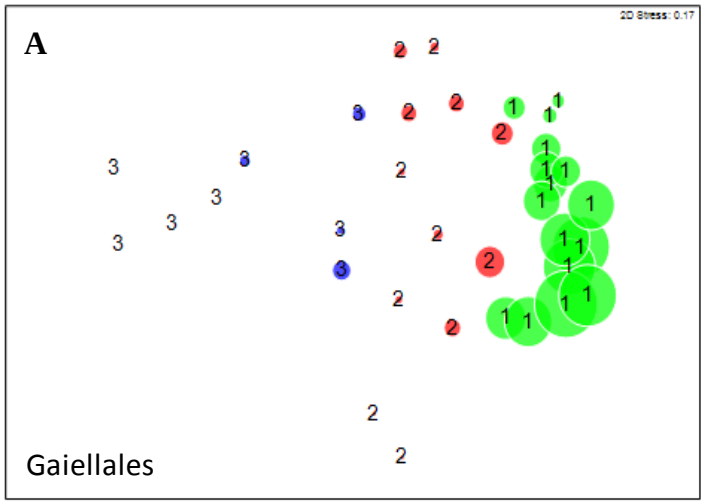
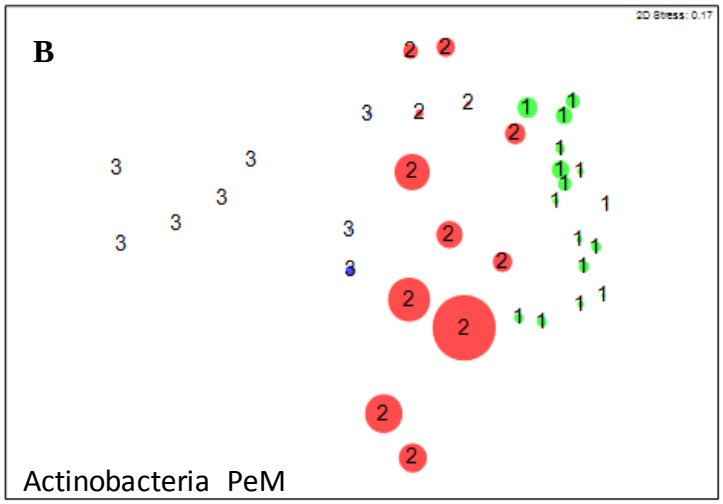
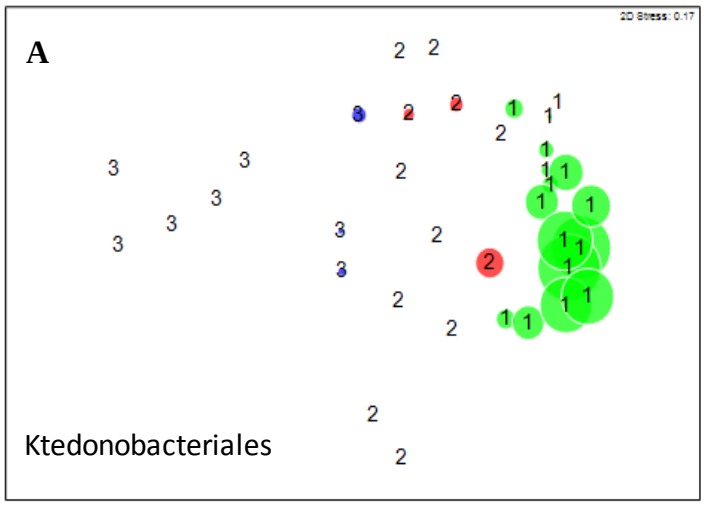
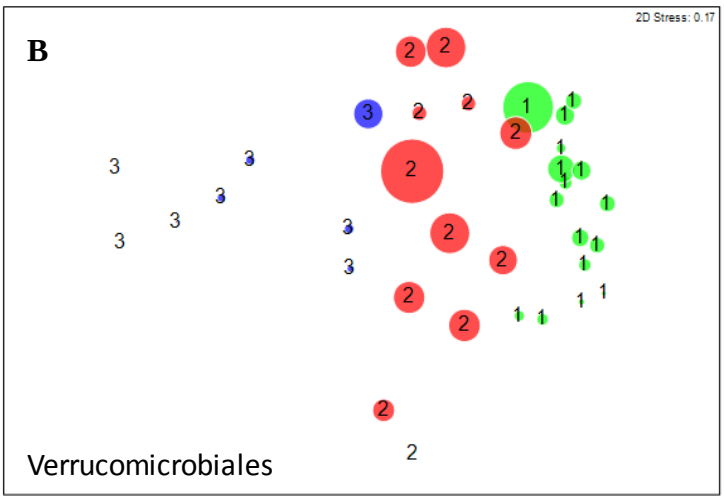


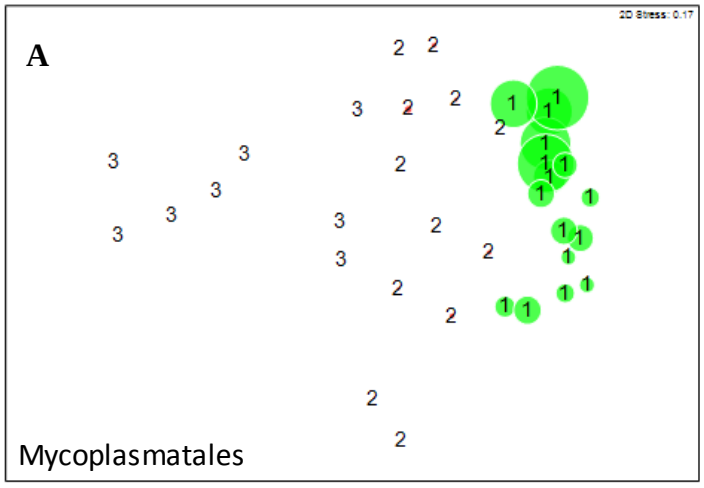
Figure S4: Triplicate water samples from cooling tower, drinking water, and river water used to assess technical variations. Despite different water types used, relative abundance of phyla was highly similar between triplicates of independently analysed samples. A high reproducibility was also detected in the rest of the taxonomic levels (class to genus level) for the same water samples (data not shown).

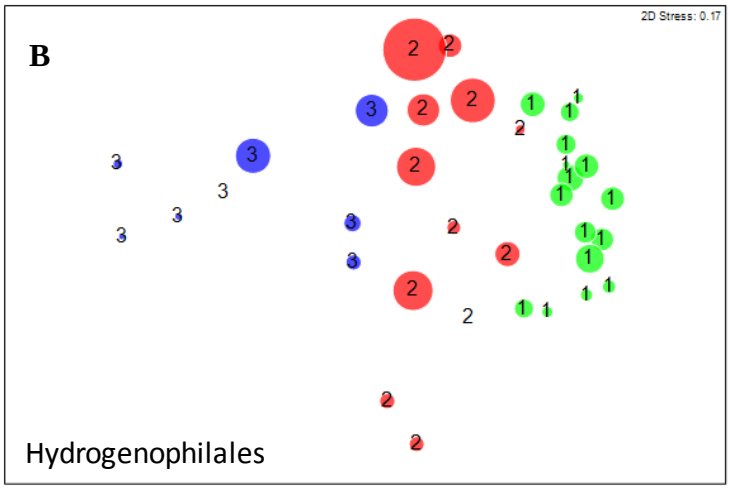

A
2D Stress: 0.17
Gaiellales


B
2D Stress: 0.17
Actinobacteria PeM


A
2D Stress: 0.17
Ktedonobacteriales


B
2D Stress: 0.17
Verrucomicrobiales


A
2D Stress: 0.17
Mycoplasmatales


B
2D Stress: 0.17
Hydrogenophilales

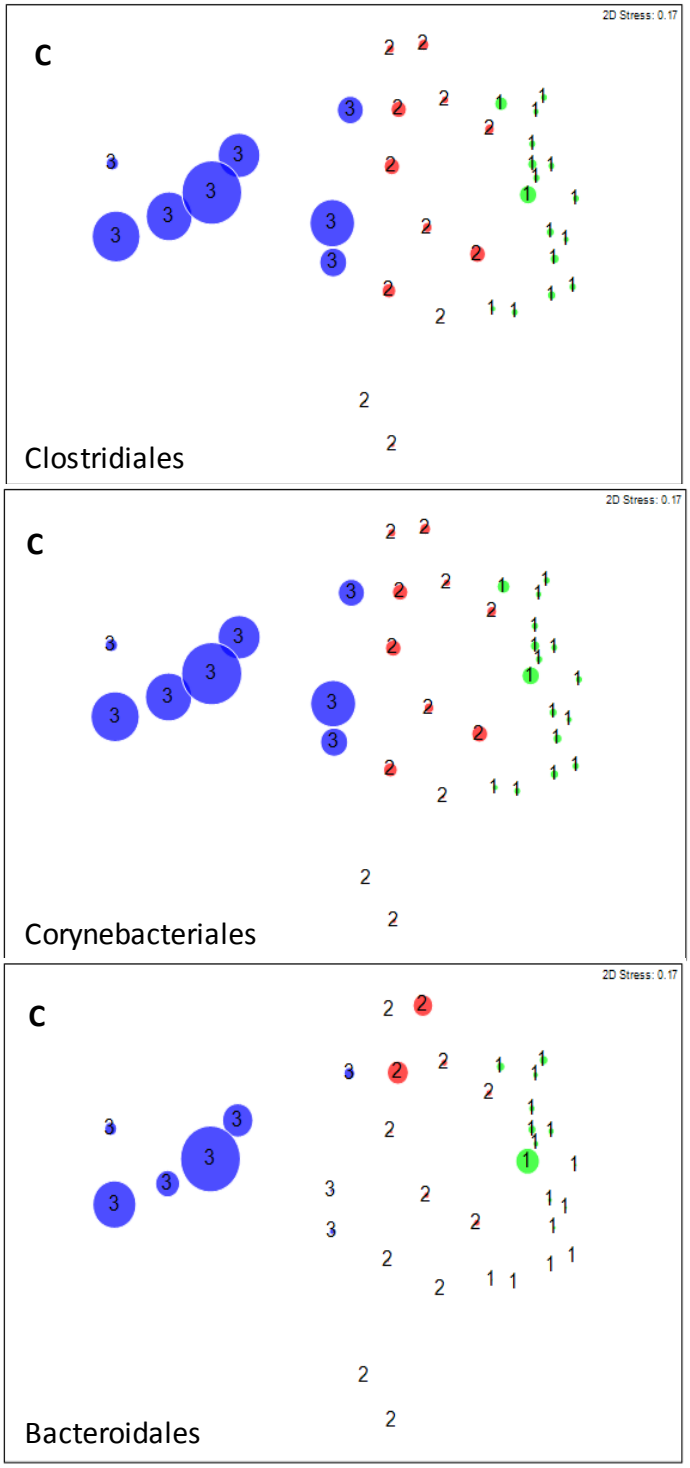


Figure S5: nMDS plots for bacterial orders enriched in pristine (A), urban (B) and Agriculture (C) sampling locations. The blue, red and green colors and their respective numbers indicate pristine, urban and agricultural sampling points, whereas the size of the bubble indicates the percentage of the sequences detected for the respective taxa at that particular site. n=36 gDNA samples.

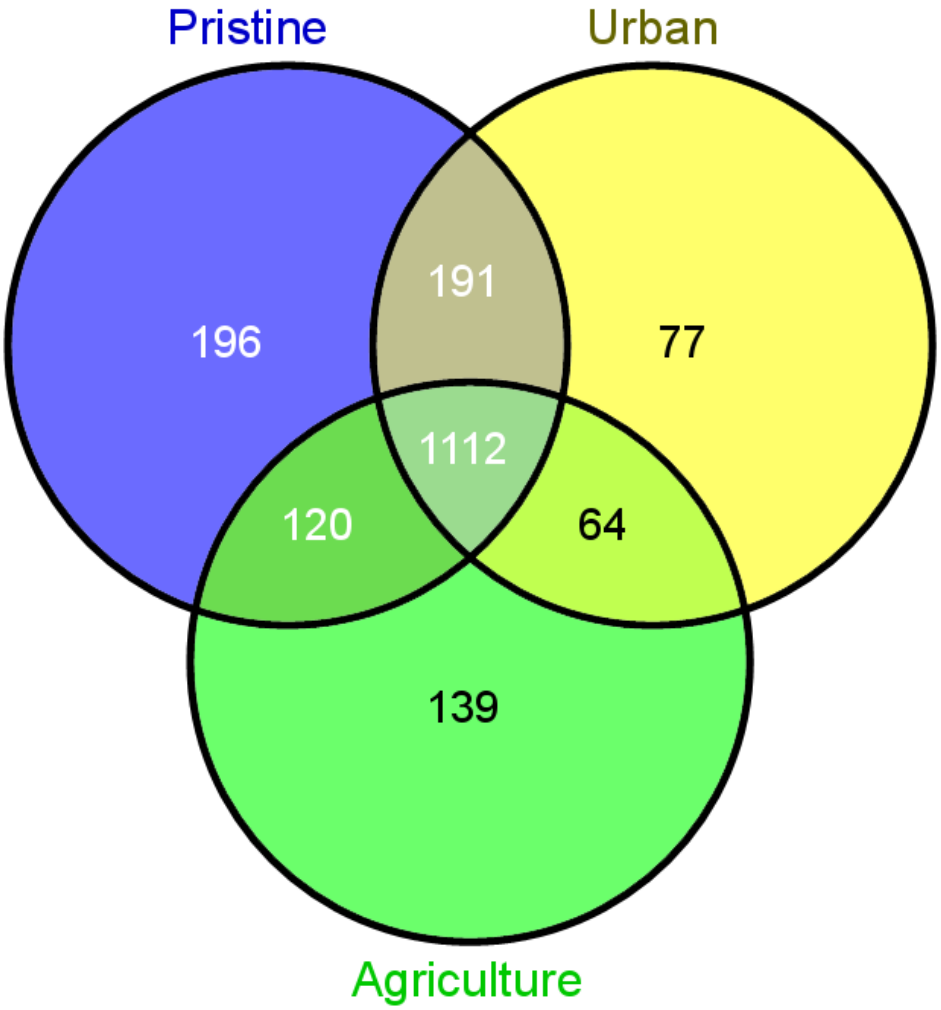


Figure S6: Significant number of the detected bacterial taxa from pristine location was present in all the investigated samples while only 9 to 13% of all identified taxa were contributed by urban and agricultural sampling sites of the investigated river system. These results suggested that pristine is the main origin of the most bacterial taxa in the investigated river water system.

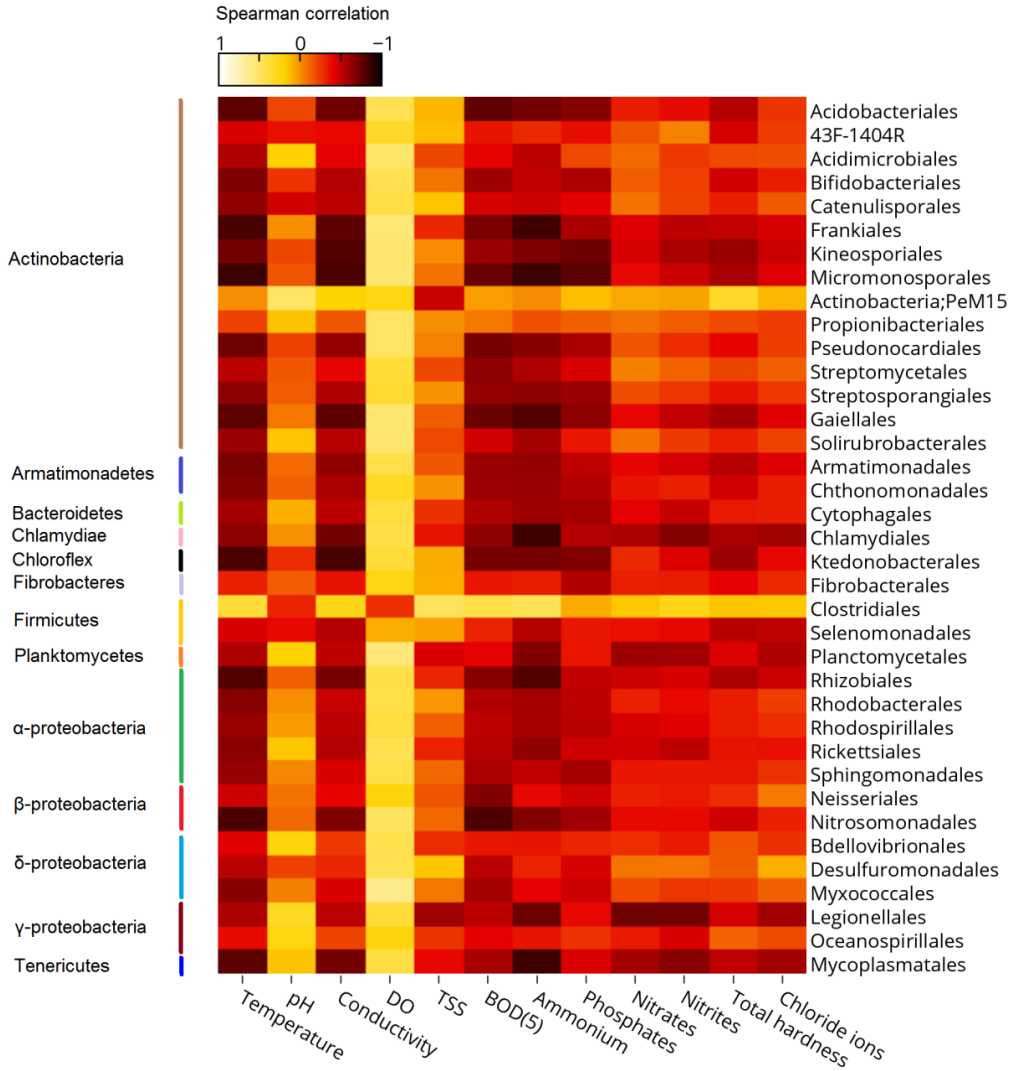


Figure S7: Heatmap of Spearman correlation between environmental variables and bacterial order detected in the analysed water samples. Colours indicate r-values of Spearman correlations between relative abundances of the 37 most abundant bacterial orders and physico-chemical variables.

Table S1. Efficiency of bacterial cells recovery from polycarbonate membrane filters.

Filter #	Cell counts in original sample /ml	Cell counts Retrieved from polycarbonate membrane /ml	Efficiency %
1	1.90E+06	1.90E+06	99.89
2	1.87E+06	1.87E+06	100.00
3	1.50E+06	1.49E+06	99.33
4	1.81E+06	1.80E+06	99.38
5	1.63E+07	1.61E+07	98.77
6	3.10E+06	3.09E+06	99.68
7	2.70E+07	2.69E+07	99.63
8	2.10E+06	2.09E+06	99.38
9	2.00E+07	2.00E+07	99.90
10	1.90E+07	1.89E+07	99.36

Table S2: One-Way ANOSIM statistics (performed at 999 permutations) based on Bray-Curtis similarity matrix derived from detected bacterioplankton communities of the investigated water system.

	Global R = 0.60 (P=0.001)	
	Pristine	Urban
Urban	0.50(P=0.001)	
Agriculture	0.83 (P=0.001)	0.51(P=0.001)

Table S3: Spearman correlation coefficients and significance values of Mantel-like RELATE tests between Bray Curtis based bacterial community distance matrices and Euclidean based abiotic distance matrices.

Variable/group of variables	Coefficient of variation (R)	*P*-value[1]	Significance[2]
Global environmental dataset[3]	0.61	<0.001	***
Temperature	0.57	<0.001	***
pH	0.05	0.163	ns
Conductivity	0.49	0.001	***
Dissolved oxygen	0.29	0.003	**
Total suspended solids	0.26	0.004	**
Biological oxygen demand	0.62	<0.001	***
Ammonium	0.69	<0.001	***
Phosphate	0.43	0.001	***
Nitrate	0.16	0.018	*
Nitrite	0.39	0.001	***
Total hardness	0.44	0.001	***
Chlorides	0.18	0.003	**

[1]Probability (*P*) values were computed using 999 permutation algorithm.

[2]Extent of significance based on the computed *P* value: ***, highly significant; **, significant; * moderate significant; ns, non-significant

[3]Calculated between bacterial community data set and whole environmental data set.

CHAPTER 4

Deep sequencing of 16S rRNA gene amplicons revealed discrete series of microbial community succession in tropical drinking water treatment plant

Mushi, D [1,2], Pereira, RPA[1], Höfle, MG[1] and Brettar I[1]

[1]Helmholtz Centre for Infection Research, Braunschweig (Germany)

[2]Sokoine University of Agriculture, P.O. Box 3038, Morogoro, Tanzania

4.0 Abstract

Microbial community dynamics in drinking water treatment plant are remarkably understudied, despite the high demand of microbiologically safe drinking water supplies. In this study, data on coherent dynamics of taxa- and microbial community shifts along the treatment barriers of drinking water treatment plant are reported. By sequencing the 16S rRNA gene amplicon at adequate depth, a high degree of microbial diversity and overrepresentation of typical freshwater genera including *Undibacterium, Novosphingobium and Cylindrospermopsis* were observed. *Undibacterium* had a considerable contribution to the abundance of the phylum *Proteobacteria* and demonstrated a remarkable ability to predict microbial diversity. Shifts in community structure were due to substantial elimination of bacterial taxa by sand filtration, and significant enrichment of rare taxa following chlorination. Coherent dynamics of taxa across treatment barriers revealed series of discrete microbial secondary successions punctuated by treatment barriers. Based on microbial community succession data, the fate of noxious bacteria in drinking water treatment plants can be tracked.

4.1 Introduction

Like many other countries worldwide (1–5), Tanzania depends on river systems as the main source of drinking water (6–8). However, the microbiological quality of the source water is often modified by stochastic episodes of contamination associated with an array of viruses, bacteria, and protozoa due to lack of appropriate catchment protection (2). Consequently, efficient treatment steps that can inactivate and eliminate noxious microbes from surface water and limit their re-growth before being distributed to the end users are inevitable. For decades, conventional drinking water treatment plants (DWTP) equipped with a train of treatment processes (e.g. aeration, coagulation/flocculation, sedimentation, sand filtration and post-chlorination) have been operating in Tanzania to treat surface water for domestic and industrial use purposes. Despite frequent occurrence of waterborne outbreaks associated with the drinking water (9), the impact of each of these treatment steps on the quality of finished water in general, and on the microbial communities in particular, is basically unknown for a

tropical DWTP such as that of Morogoro City in Tanzania. Hence, a thorough understanding of the microbial ecology of tropical DWTP is highly needed in order to improve microbial safety of drinking water and operational monitoring practices.

A number of scientific studies have investigated the I6S rRNA-based microbial community dynamics across multistep DWTP during the last ten years (5,10–15). Although these studies identified filtration (10,11) and chlorination (12,13,16) as key treatment steps shaping microbial communities in DWTP, a broad perspective of complex microbial communities in DWTP is still lacking because these studies employed 16S rRNA-based detection methods that offer insufficient sequencing depth (5). On the other hand, none of these studies have demonstrated the coherent dynamics of taxa across treatment barriers in the DWTP despite the fact that changing water quality during the course of treatment could selectively enrich different microbial taxa (11) including those with public health concern (17).

High-throughput 16S rRNA gene sequencing using Illumina MiSeq platform provides an opportunity to: exhaustively profile the change of complex microbial communities throughout a drinking water treatment plant, understand how each treatment step influences the microbial communities, detect bacterial taxa regrowth, and allow understanding of the microbiological safety of the finished water (11). Despite the acknowledged high resolution of deep sequencing technology (18) especially in resolving rare biosphere such as pathogenic bacteria (17), their application to explore complex microbial communities of DWTP is far more in the infancy compared to what have been done in wastewater (19–21), soil (22), marine (23), natural freshwater bodies (24,25) and human microbiome (26). To this end, we applied deep sequencing using Illumina MiSeq platform to comprehensively analyze the complex microbial communities of a full scale tropical DWTP that produces drinking water for 85% of the population (> 300,000) residing in Morogoro City, Tanzania. The tropical DWTP treats river water (stored in Mindu reservoir) through aeration, coagulation/flocculation/pre-chlorination, sedimentation and post-chlorination, respectively before being distributed to end users. Specifically, we characterized microbial communities of tropical DWTP by deep sequencing of V4 and V5 hypervariable regions of the 16S rRNA gene using Illumina MiSeq platform, and

determined the treatment barriers significantly contributing to the variability of the drinking water treatment plant microbiome. Taxa coherence dynamics across treatment barriers were investigated to reveal the microbial succession in the drinking water treatment plant. In addition, the influence of microbial diversity on the dynamics of dominant taxa was elucidated. In this study, we document for the first time, the presence of microbial succession in the drinking water treatment plant, and highlight the importance of geographical location of source water on the microbial composition found in the drinking water treatment plant and the finished drinking water.

4.2 Materials and Methods

4.2.1 Location and characteristics of drinking water treatment plant

Drinking water treatment plant was situated at 6°83´S and 37°63´E in the suburb of Morogoro City, Tanzania. The plant treated water pumped from the Mindu reservoir with an area of about 400 hectares and a maximum depth of approximately 11 m (27).The reservoir received water from the upper Ngerengere watershed that originates from the North-Western part of Uluguru Mountain (see Figure 2B in Chapter I). Apart from serving as a reservoir for drinking water supply, fishing is also performed in the reservoir. The Drinking water treatment plant produces approximately 24 million litres of water per day, which serves 85% of the Morogoro municipal population (> 300,000 people) (http://moruwasa.co.tz/home/index.php?).

4.2.2 Water treatment processes

After screening the reservoir water to remove debris originating from water plants, water is subjected to aeration cascades to enhance the air-water transfer and for the removal of volatile components such as methane, hydrogen sulphide and ammonia (Fig. 1). Water free from odour gases is directed to the reaction compartment where coagulation and flocculation processes are performed. In contrast to other drinking water treatment plants, pre-chlorination is integrated in the reaction compartment as an additional treatment (Fig. 1). To remove suspended particles, water is subjected to sedimentation and sand filtration respectively before

being re-chlorinated. To maintain the microbiological safety of the finished water, sand filters are backwashed regularly as described in detail by Albers *et al*. (28).

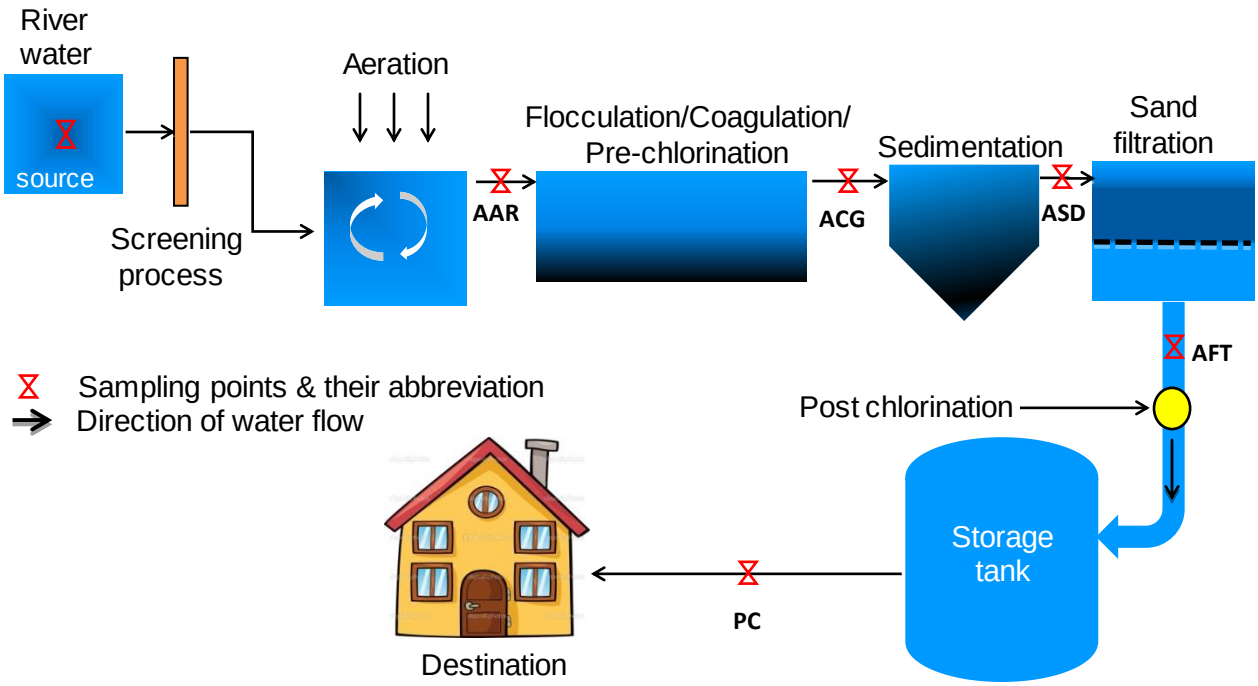


Figure 1: Schematic representation of drinking water treatment processes of the water utility that produces drinking water for the Morogoro City in Tanzania. The main sampling points are shown by a red sign and their corresponding abbreviations. The sampling sites: SW, source water; AAR, after aeration; ACG; after flocculation/coagulation/prechlorination; ASD, after sedimentation; AFT, after sand filtration; PC, post chlorination.

4.2.3 Synoptic water sampling and pre-treatment

In order to capture only the influence of the treatment barriers on microbial communities and composition, water was sampled in dry period months spanning July to August, 2014 from raw water, and from effluents of each treatment step. These include triplicate water samples (4 l each) from Mindu reservoir, and from effluent of each treatment step resulting to a total of 18 samples. Water was sampled aseptically using 1 l wide-mouthed sterile plastic bottles (Thermo Scientific™ Nalgene™, Neubrecht, USA) and immediately placed into a dark, ice-cooled (4°C)

box before being transported to Sokoine University of Agriculture (Morogoro, Tanzania) laboratory for pre-processing.

Bacterial cells were recovered from sampled water and stored on FTA™ classic card (GE Healthcare UK Limited, Buckinghamshire, UK) within 6 to 9 hours of collection of the first sample. Briefly, sampled water was filtered onto a 0.2-µm-pore-size polycarbonate filter (47 mm diameter; Nuclepore; Whatman, Maidstone, UK). After filtration, the filter was placed to a 47 mm sterile petri dish and kept wet with Milli-Q water. Bacterial cells were scraped off the filter surface as described elsewhere (29). Suspended cells in Milli-Q water were pipetted from the petri dish and transferred to the drawn circle area of the FTA™ classic card. The filter was rinsed twice with Milli-Q water, the rinsate was also transferred to the FTA™ classic card. Efficiency of retrieving bacterial cells from polycarbonate filters ranged from 99 to 100% (see Supplementary Table S1 and associated information therein), suggesting that cells were retrieved quantitatively from the polycarbonate filters. FTA™ Classic cards containing the bacterial consortia were allowed to dry overnight on sterile silica gel before being placed into a labeled multi-barrier pouch (GE Healthcare Europe, Freiburg, Germany) to maintain sample integrity during storage. Multi-barrier pouch contained samples were sealed and transported to Helmholtz Centre for Infection Research (HZI), Braunschweig, Germany for further processing. It should be highlighted that the FTA™ classic card was found to be the best tool for storing genomic DNA (gDNA) at room temperature for an extended period of time (> 2 months) without degradation following an experiment that compared several documented non-freezing methods for DNA preservation (see Supplementary Fig. S1 and associated information therein).

4.2.4 Genomic DNA (gDNA) extraction

A total of 6 FTA punches (each with 6 mm diameter) containing DNA of the applied bacterial cells were punched out aseptically from FTA™ Classic cards using a metal single hole puncher and immediately placed into a sterile water bead tube (Mo Bio laboratories inc., Carlsbad, CA, USA). To avoid cross contamination of samples, the metal puncher was rinsed with 70% ethanol and flamed for 10s before being applied to a subsequent sample. gDNA was extracted from FTA punches using UltraClean™ PowerWater DNA isolation kit (Mo Bio laboratories inc., Carlsbad,

CA, USA), following the manufacturer's instructions. The quantity of the extracted gDNA was gauged spectrophotometrically using NanoDrop 2000c spectrophotometer (NanoDrop Technologies, Wilmington, DE, USA) followed by quality analysis using 260/280 nm and 260/230 nm absorption ratios and agarose gel electrophoresis, respectively prior to storage at -20 ^{o}C for later use.

4.2.5 Library construction and deep sequencing of 16S rRNA gene amplicons

16S rRNA gene libraries were generated as described in Camarinha-Silva *et al*., (30) employing COM1F (5´-CAGCAGCCGCGGTAATAC-3`) and COM2R (5'-CCGTCAATTCCTTTGAGTTT-3') universal primers that amplifies approximately 410 bp of V4 and V5 hypervariable regions of most bacterial 16S rRNA gene (31). Incorporated to the COM1F for each sample is the distinctive 6-base error correcting barcode and 2-base CA linker to allow for allocation of each sequence to the appropriate sample, and to avoid amplification bias (32,33), respectively. On the other hand, COM2R was fused with unique 6-base index to allow multiplexing of the samples. Both primer pairs contained appropriate adapters at the 5´ ends to permit sequencing on the Illumina MiSeq platform (30,34). An amplification of 50 µl reaction mixture contained: 100 µM deoxynucleoside triphosphate (Bioline, Luckenwalde, Germany), 0.4 mM $MgCl_2$, 1× PCR reaction buffer, 0.03 U HotStarTaq Polymerase (Qiagen, Hilden, Germany), 0.4 µM of each primer and 2ng of gDNA template. PCR was performed on a Biorad Thermo cycler 96-well iCycler with an initial denaturation cycle of 95^{o}C for 15 min, followed by 30 cycles at 95^{o}C for 1 min, 55^{o}C for 40s, and 72 ^{o}C for 40s; a final extension of 10 min at 72^{o}C was added to ensure complete amplification. Two µl from the first PCR reaction mixture were used as template in a second PCR reaction performed under the same conditions as for the first PCR, except that 10 cycles, and PCR primers designed to integrate the sequence of the specific Illumina multiplexing sequencing primers and index primers were employed (30). Template free controls (using water instead of DNA) were performed in the two steps PCR procedure in parallel with the considered samples and, as they did not result in amplification, were not further considered in downstream amplicon processing (30).

To isolate 16S rRNA amplicons with approximately 410 base pair, PCR products were run on 2% agarose gel electrophoresis pre-stained with GelRed Nucleic Acid Gel Stain (Biotium, Hayward, CA, USA). Desired 16S rRNA bands on agarose gel were identified with the aid of blue light transilluminator. Bands were carefully excised and amplicons were extracted from the agarose gel according to the condition described in van Dijk *et al*., (35) using QIAquick Gel Extraction Kit (Qiagen, Hilden, Germany). The extracted amplicons were then quantified with Quant-iT PicoGreen dsDNA reagent (Life Technologies, Oregon, USA). Library tagged with unique index was prepared by pooling equimolar ratios of amplicons (≈30 ng of each sample) from an appropriate number of samples. Prior to deep sequencing, normalized libraries were purified using QIAGEN MinElute® PCR purification Kit (Qiagen, Hilden, Germany), quality determined, and quantified using Agilent BioAnalyzer platform. Sequencing of the quantified libraries was performed at genome analytics facility (GMAK) of the Helmholtz Centre for Infection Research (HZI), Braunschweig, Germany using the Illumina® MiSeq platform.

4.2.6 16S rRNA gene sequence data analysis

16S rRNA sequences generated by Illumina MiSeq platform were processed using Mothur software package version 1.27.0 (36). Briefly, the 933 755 raw sequences obtained from 18 samples of genomic DNA sequenced by Illumina MiSeq platform were trimmed to remove primer and barcode sequences, and to eliminate poor quality sequences such as sequences shorter than 200 bp, those that contained ambiguous base greater than 2%, those that comprised 2% nucleotide homopolymers, and those that had a quality score less than 40 in a sliding window of 50 nucleotides over the sequence length. Residual clean sequences were screened for chimeras using UCHIME algorithm (37) in Mothur and any sequences suspected to be chimeras were excluded from downstream analysis. High quality sequences (916 945) that persisted a train of stringent quality control were aligned against the SILVA data base Incremental Aligner (38). Sequences were assigned to different operational taxonomic units (OTUs) at 97% sequence similarity by using Mothur and the RDP taxonomic data base release 9 (39).

4.2.7 Statistical assays

Explicet software (40) was used to display pie chart, and to generate alpha diversity metrics including Good's coverage, observed richness, estimated richness (Chao 1), ACE, Shannon (H') and Simpson (1-λ) diversity indices after standardization of 16S rRNA sequences to 14, 089. Microbial community structure was analysed by PRIMER version 7.0.6 software (PRIMER-E, Playmouth, UK) based on Bray-Curtis similarity algorithm, and visualised by canonical analysis of principal coordinates (CAP) plot. Another algorithm, such as Sørensen similarity was also applied and produced equivalent outcome for the given data set (See supplementary Fig S2). Non-parametric similarity profile (SIMPROF type 3) analysis that take into account relative abundance and index of association was used to detect statistically discrete subsets of taxa which responded to treatment barriers in a coherent manner (41,42). A heatmap for normalized relative abundance (forth-root transformed) of genera was generated according to the method described by Parks and Beiko (43). Correlation analyses were done with the Statistical Package for Social Sciences, version 16.0 (SPSS Inc. Chicago, III.).

4.3 Results

4.3.1 Adequacy of the sequencing depth and community diversity.

From the sequenced fragment spanning the V4 and V5 hypervariable regions of the 16S rRNA gene (31), 916,945 high quality Illumina MiSeq sequences (average length of about 410 bp) were recovered across 18 samples of genomic DNA with an average of 50,941 sequences per sample (range, 15 427 - 68 767). Following sequence assignment to different operational taxonomic units (OTUs) at 97% similarity, a total of 15,762 OTUs were obtained. Of these, 99.9% belonged to the bacterial domain while the remaining fraction represented taxa affiliated to archaeal domain. The calculated rarefaction curves (Fig. 1A) based on the number of OTUs and their corresponding sequences for each sample approached the saturation point, indicating that the sequencing depth per sample was appropriate in resolving the optimal number of OTUs needed for representation of the entire microbial diversity of the investigated DWTP. This observation was further supported by Good's coverage values that averaged to 96.3% ± 0.03% (mean ± SD, range = 92-100%) per sample when individual samples were

rarefied to the same number of sequences (14,089 sequences/sample) to normalize the sampling effort (Fig. 1B).

RDP classification performed on the high quality sequences suggested that tropical DWTP comprised diverse microbial consortia as proven by sequences detected from >26 phyla. However, >91.9% of the overall detected 16S rRNA sequences were diverse representatives of only five phyla (Fig 2). Phylum *Proteobacteria* was the most predominant and ubiquitous in the tropical DWTP comprising 59.36% of the overall detected sequences. Within the phylum *Proteobacteria, Betaproteobacteria* was the most dominant (37.1%) followed by *Alphaproteobacteria* (13.3%), *Gammaproteobacteria* (7.3%) and *Deltaproteobacteria* (1.53%). *Betaproteobacteria* and *Alphaproteobacteria* were highly dominated by the genera *Undibacterium* (24.95%) and *Novosphingobium* (5.75%), respectively. *Bacteroidetes* (13.53%) constituted the most abundant phyla beyond *Proteobacteria* followed by *Cyanobacteria* (8.89%, of which 5.01% was the genus *Cylindrospermopsis*), *Actinobacteria* (5.64%) and *Planctomycetes* (4.44%). A similar picture was observed when the sequence distribution for each phylum across all samples of the studied DWTP was considered (supplementary Fig. S2).

The discrete bacterial genera comprised in the identified phyla displayed various range of occurrences across the sampling points that generally reflected their differential response towards the treatment barriers, and stochastic processes happening in the source water. From 1269 genera detected in this study, 14 (1.1%) were present in all sampling points throughout the investigation period. Interestingly, the three most abundant genera (*Undibacterium*, *Novophingobium*, *Cylindrospemopsis*) were among of those with the highest prevalence (100%). About 921 (72.6%) genera displayed a prevalence ranging between 10-90%. On the other hand, 334 (26%) genera were detected only once (prevalence <10%) in the whole data set.

Rank abundance curves (supplementary Fig. S3) showed that between 98% and 99% of total genera detected in individual samples had a relative abundance below 1% with the majority occurring in <100% of the investigated samples as proven by the strong relationship between average relative abundance and prevalence (supplementary Fig. S4). It is very difficult to believe that such a huge percent of genera can be due to sequencing errors. Data suggested that some

genera such as *Acinetobacter* and *Burkholderia* had on average a relative abundance of 0.07% ± 0.001 (mean ± SD) and (0.73% ± 0.03), respectively and were detected continuously in all considered samples in this investigation (prevalence of 100%) including replicates. We observed many genera with low abundance (<1%) and high prevalence (10-90%) including *Sphingopyxis* (0.03 ± 0.0004), *Rhizobium* (0.04 ± 0.001) and *Synechococcus* (0.25 ± 0.003) among others (see also Fig. S4). On the other hand, enrichment of rare taxa during the water treatment process was evident (Table 1).

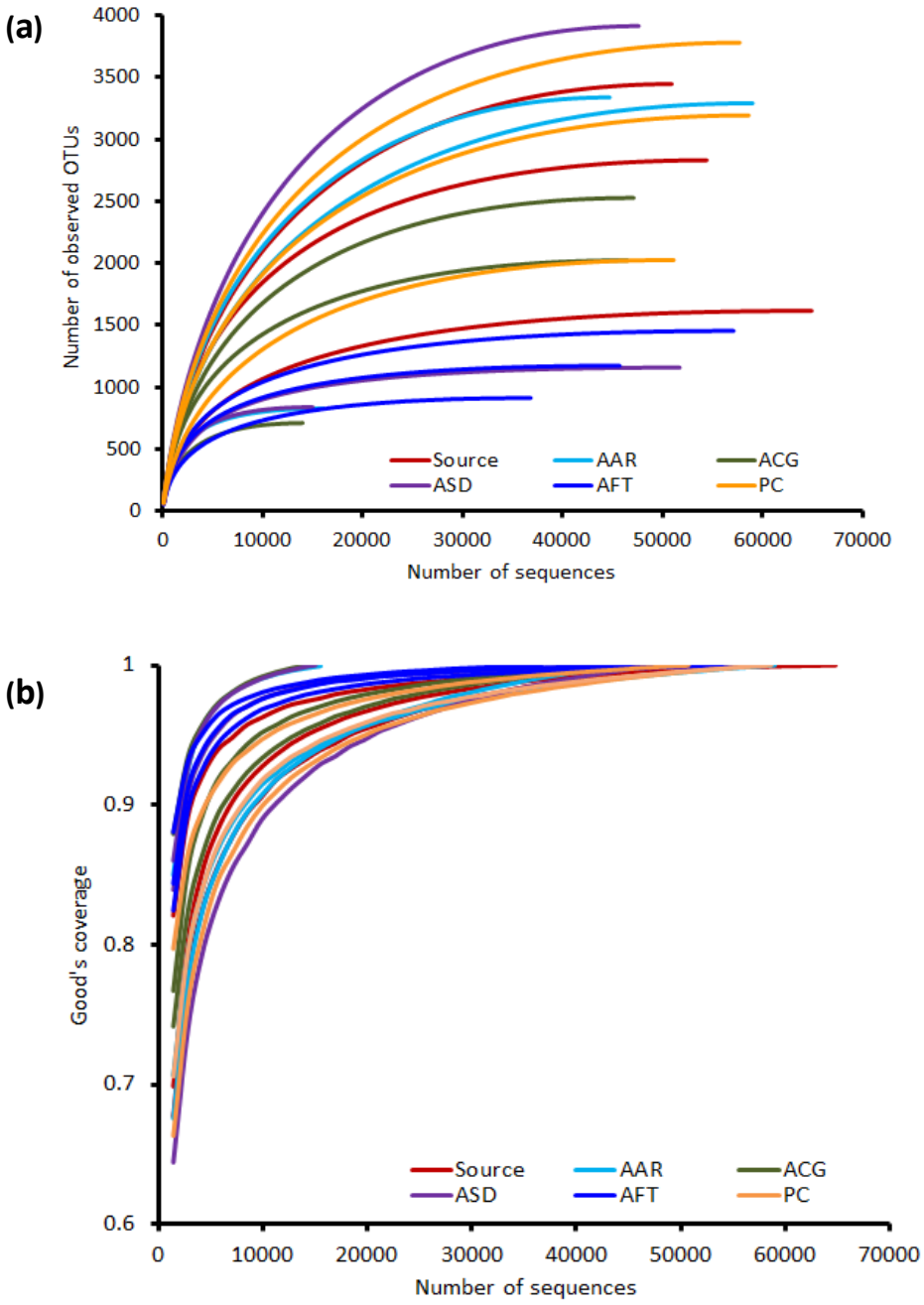


Figure 1. OTUs (a) and Good's coverage (b) collector's curves articulating the adequacy of the sequencing efforts for the complete representation of bacteria communities that were studied. The curves generated using the two alpha diversity metrics - for the considered samples - reached asymptotes, indicating that all water samples were sequenced reasonably well to be considered representative of the bacteria communities that were analyzed. For each plot, curves of the same color exemplified samples from the same sampling point (n=3). Abbreviations: AAR, aeration effluent; ACG, coagulation/flocculation/pre-chlorination effluent; ASD, sedimentation effluent; AFT, sand filtration effluent; PC, post chlorination effluent.

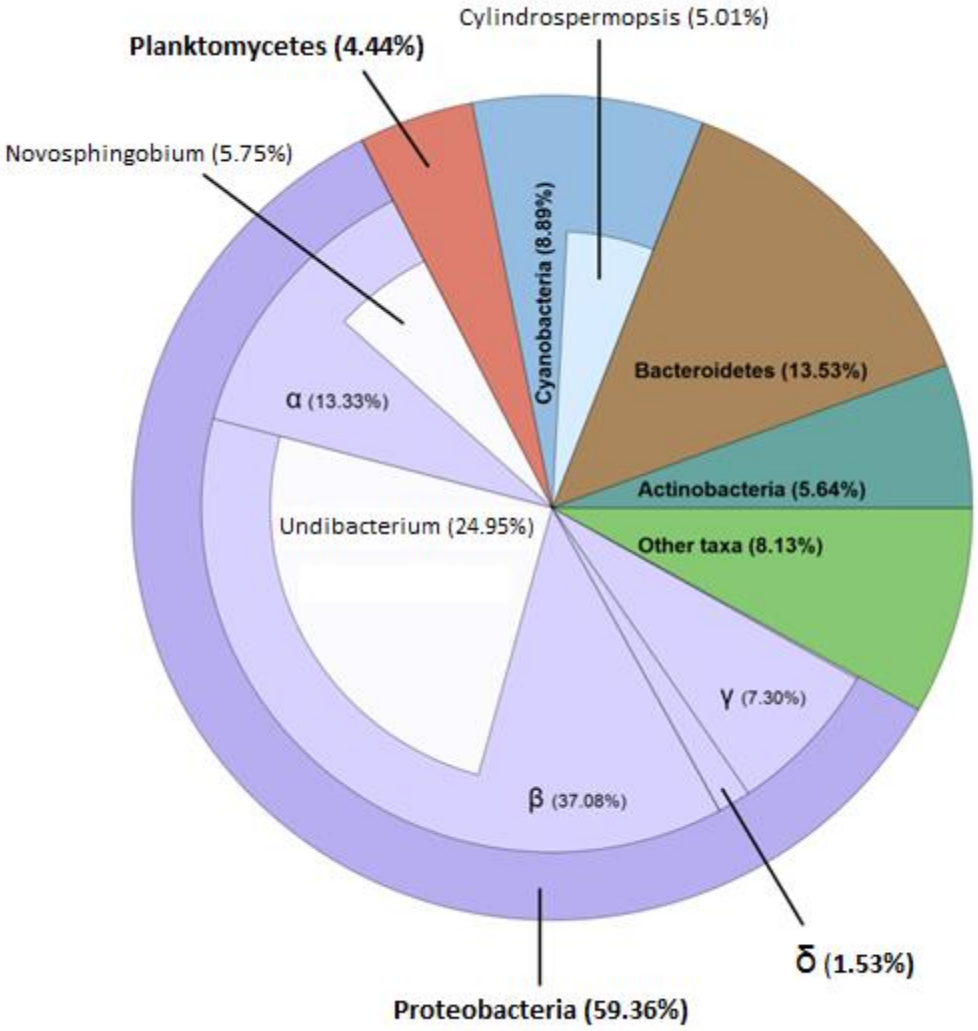


Figure 2: The most abundant phyla and genera observed across all 6 sampling points considered for this investigation. Five most abundant phyla contributing 91.9% of the detected sequences (916,945) from 18 samples (3 triplicates per sampling point) were *Proteobacteria* (59.36%), *Bacteroidetes* (13.53%), *Cyanobacteria* (8.89%), *Actinobacteria* (5.64%) and *Planctomycetes* (4.44%), whereas 21 minor phyla each exhibiting <<4% of the total sequences were collapsed into a group denoted as ``Other taxa´´. The most dominant genera were *Undibacterium*, *Novosphingobium* and *Cylindrospermopsis*, respectively that are presented as the smallest nested wedges in their respective phylum.

Fifty three genera had a relative abundance >1% in at least one of the sequenced samples (Fig. 3). These assemblages comprised of ubiquitous freshwater planktonic-(e.g. *Cylindrospermopsis*)

(44) and soil (e.g. *Acidobacteri*a) (45) bacteria, and those known for biofilm formation (e.g. *Sphingomonas*) (46). Of note, genera belonging to the phylum *Proteobacteria* were widely represented in these assemblages compared to those from other phyla. Genus *Undibacterium* dominated 4 of 6 samples investigated in this study and their abundance explained 73% of the variation of relative abundance of phylum *Proteobacteria* (Supplementary Fig. S5). The most diverse samples, aeration and post-chlorination effluents were not dominated by *Undibacterium* instead by genus *Cylindrospermopsis* (Mean ± SD; 8.5% ± 1.5%) and uncultured lineages of the family *Cytophagaceae* (8.47% ± 0.1%), respectively, suggesting that high microbial diversity may have an impact on the ecology of *Undibacterium*. Interestingly, sand filtration effluent contained higher relative abundance of *Undibacterium (*66.6% ± 0.24%*) and Novophingobium* (23.0% ± 0.19%) than any of the genera within and across samples (Fig. 3). Such episodic increase in abundance happened after 84% of the genera contained in the source water had been eliminated from water by sand filtration as depicted by the data from sequenced 16S rRNA-gene amplicons (Fig. 3).

The universal 16S rRNA primers applied in this study could allow the detection of low abundant genera that comprised opportunistic human pathogens in the post-chlorination effluent. These genera included, *Legionella* (abundance range*,* 0.74 - 1.63%) *Sphingomonas* (2.17 - 3.27%*), Mycobacterium* (0.04 - 0.06%), *Rickettsia* (0.01 - 0.07%), *Pseudomonas (*0.07 - 0.15%*)*, *Aeromonas* (<0.01%), *Burkholderia* (<0.01 - 0.05%*), Acinetobacter* (0.01 - 0.04%) *and Elizabethkingia (<0.1%)*. In the same vein, taxa known to be reservoirs of antibiotic resistance such as *Bosea, Afipia*, *Sphingomonadaceae*, *Acinetobacter* and *Pseudomonas* (47–50) were also detected in the finished water. On the other hand, the toxin producing *Cyanobacteria*, *Cylindrospermopsis* was detected throughout the investigation period with abundance peak in the sedimentation effluent (Fig.3). This taxon is known for production of cylindrospermopsin toxin that is threat to public health (51). Although it is not clear whether their abundance detected in post-chlorination can cause an adverse effect to human health, their presence in water designed for drinking purposes deserve further investigation.

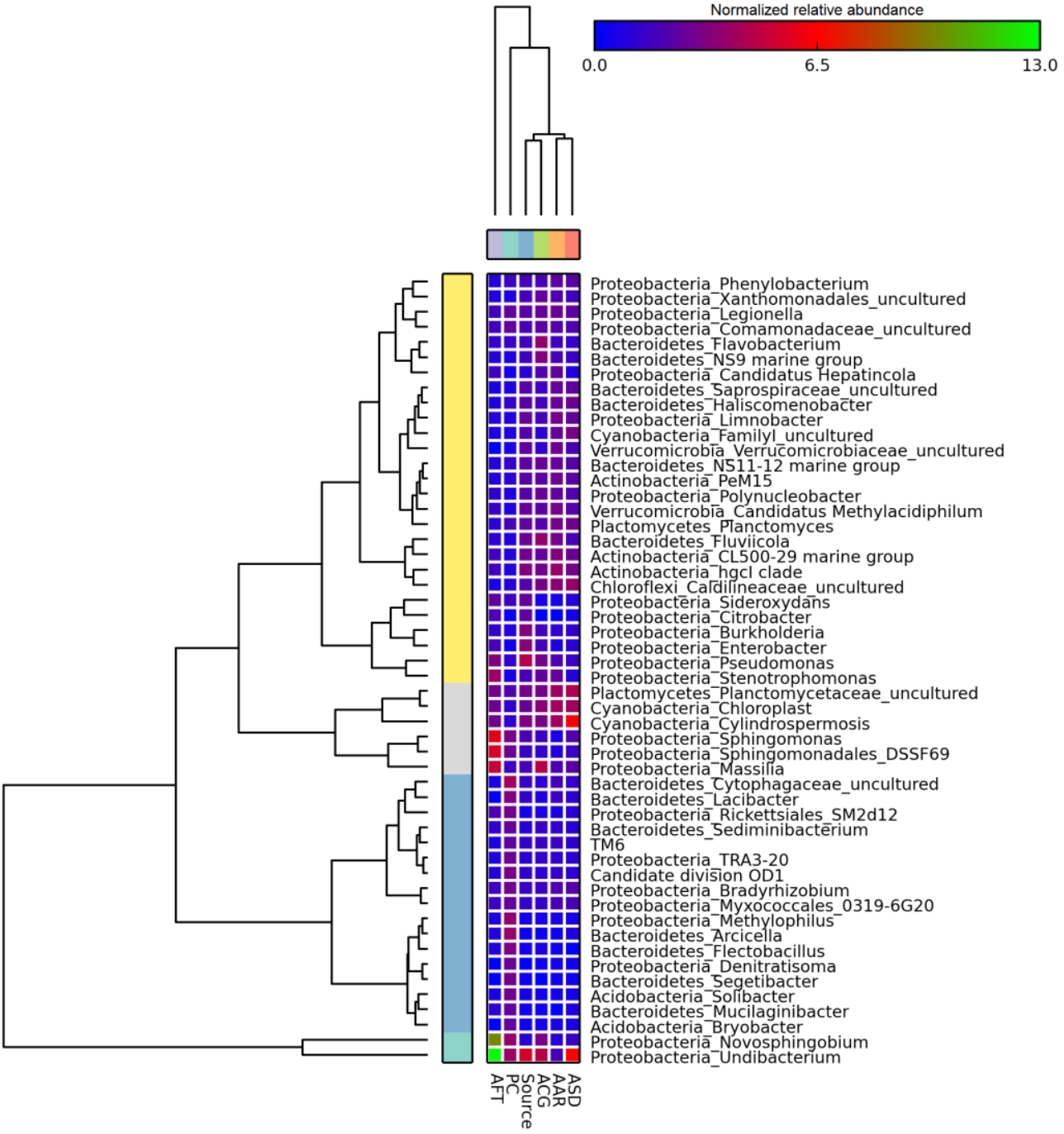


Figure 3. Heatmap showing all 53 genera (name in Phylum/genus format) with >1% average relative abundance in at least one of the sequenced samples. Shown in the upper and left of the heatmap are dendograms for water sample, and for bacteria at genus level, respectively determined using average neighbor clustering algorithm (UPGMA) and relative abundance data set. Abundance values were forth-root normalized and presented as color intensity. Abbreviations: AAR = aeration effluent; ACG = coagulation/flocculation/pre-chlorination effluent; ASD= sedimentation effluent; AFT=sand filtration effluent; PC=post chlorination effluent.

4.3.2 Effect of treatment processes on microbial diversity

Drinking water treatment processes had a pronounced impact on microbial community diversity of an individual sampling point as shown by alpha diversity metrics following OTUs rarefaction to 14,089 sequences per sample to standardize sampling effort. Alpha diversity metrics of both source water and aeration effluents were in similar ranges (Table S2). However, alpha diversity showed a drastic drop from aeration effluent to sand filtration effluent as a result of substantial removal of OTUs by sand filtration and high dominance of *Undibacterium* and *Novosphigobum* (accounting for 89% of the sequences in sand filtration effluent) (Fig. 3). The decreased diversity, however, regained more or less to original level in the post-chlorination effluent, suggesting the occurrence of microbial regrowth after chlorination. This was supported by the fact that microbial composition in post-chlorination effluent was completely different from the rest of the samples considered in this study (Fig. 3).

4.3.3 Effect of treatment processes on microbial community structure

Similarities and differences between the 16S rRNA gene-based microbial communities of source water and effluent of different treatment barriers were realized by Bray-Curtis similarity matrix, a measure of community similarity based on standardized relative abundance of genera and visualized on canonical analysis of principal coordinationates (CAP) (Fig. 4). In consistent to the diversity statistics, the differential response of microbial communities to the perturbation posed by drinking water treatment barriers was apparent. CAP analysis suggested that source water, aeration, coagulation/flocculation/pre-chlorination and sedimentation effluents had similar microbial communities but significantly different from that of sand filtration-, and post-chlorination effluents (Global ANOSIM, R = 0.5, P = 0.003, 999 permutation). Furthermore, microbial communities from sand filtration were very different from that of post-chlorination (ANOSIM, R = 1, P = 0.01, 999 permutations), indicating that sand filtration and post-chlorination are key treatment processes responsible for the shift in microbial communities of tropical DWTP. Application of the Sørensen algorithm that uses presence/absence for the same data set generated comparable outcome. Furthermore, the qualitative variation of microbial community structure supported the low number of core taxa observed in this investigation. Similarity percentage (SIMPER) analysis could reveal the most abundant genera in the data set

that were responsible for the major changes in microbial community structure. *Undibacterium* (contribution, 34.1%), *Novosphingobium* (13.8%), *Cylindrospermopsis* (6.2%) and Chloroplasts (2.9%) were found to be the most contributing taxa to the dissimilarity (average of 82.6%) of microbial communities between sand filtration effluent samples and all samples prior to sand filtration whereas, *Undibacterium* (11.6%), *Cylindrospermopsis* (6.0%), uncultured lineages of family *Cytophagaceae* (4.8%) and *Arcicella* (4.8%) were the most contributing taxa to the observed dissimilarity (average of 87.6%) between post-chlorination and all samples prior to sand filtration. Dissimilarity (83.2%) between sand filtration and post-chlorination was mostly contributed by *Undibacterium* (35.8%), *Novosphingobium* (12.4%), uncultured lineages of *Cytophagaceae* (5.1%) and *Arcicella* (4.1%).

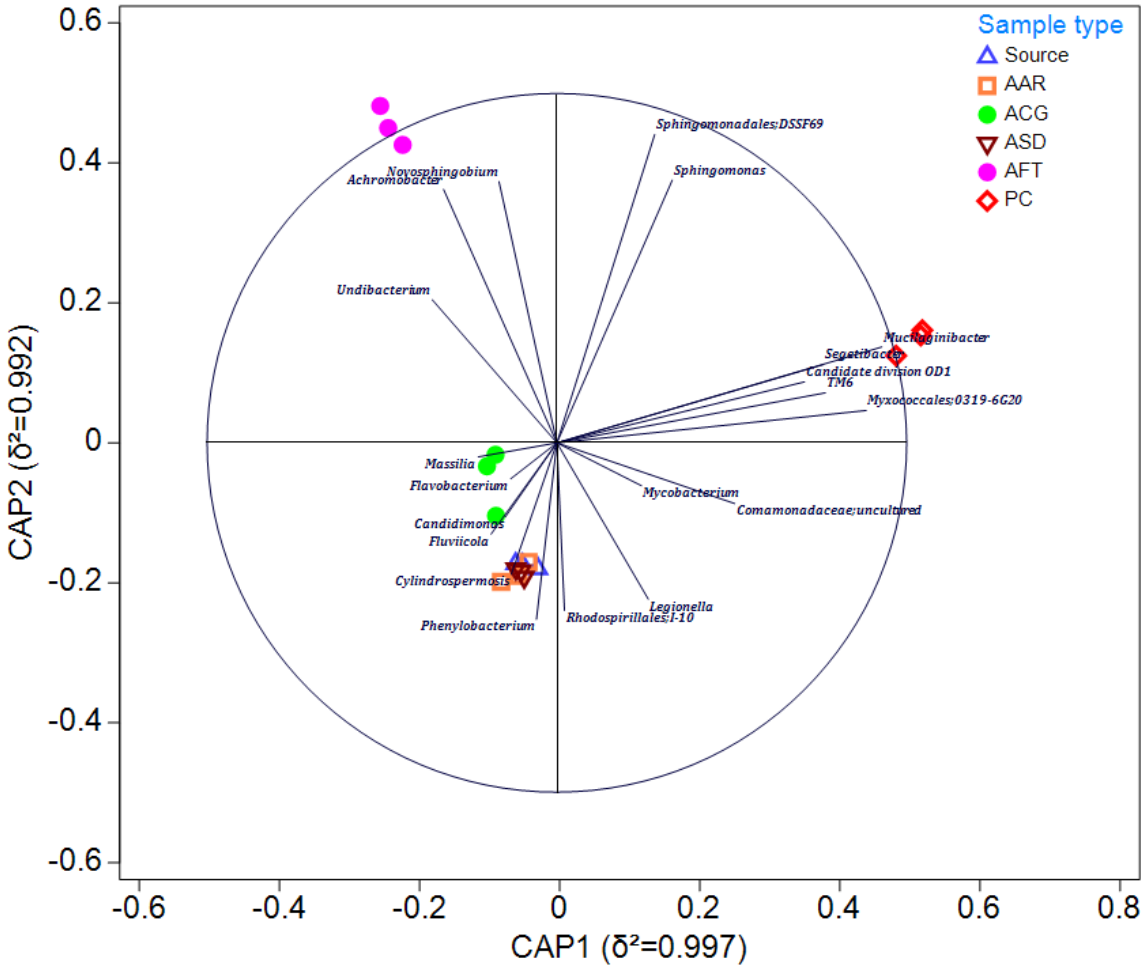


Figure 4: Microbial community structure and vectors of those taxa sharing similar trend of elevated relative abundance in drinking water treatment plant. For abbreviation, see Figure 1.

To understand which taxa were overrepresented in each of the identified microbial communities, Pearson correlation was applied on the data set that generated CAP plot. Vectors with statistically significant correlation ($r>0.3$; $p<0.05$) for those taxa with elevated relative abundance after each treatment barriers were overlaid on CAP ordination plot. Positively correlated vectors of different taxa pointed to similar direction and conformed to the pattern of microbial community structure displayed previously in this study (Fig. 4), suggesting the existence of microbial secondary succession in the effluents of treatment barriers. Microbial community succession was further explored by examining patterns of taxa with coherent dynamics using SIMPROF type 3 coupled with index of association as emphasized below.

4.3.4 Microbial secondary succession in drinking water treatment plant.

In order to evaluate response of microbiota across treatment barriers, and to identify patterns of coherently co-occurring genera, a series of type 3 SIMPROF analyses that take into account relative abundance of genera and an index of association were performed. Despite the fact that genera detected in the drinking water treatment plant originated from the same source water, their responses against treatment barriers resulted in discrete trajectories. Type 3 SIMPROF test revealed a dendogram with six significantly distinct clusters of genera that covaried coherently (at $P=0.02$) across the sampling points (Fig 5a). Members of each cluster were not taxonomically related, as taxa from distinct phyla were coherently co-occurred. The relative abundance (average of triplicate samples) of each pattern across the sampling points could confirm the presence of discrete series of secondary successions in tropical drinking water treatment plant that were punctuated by treatment barriers (5b). Coherent group **B** contained those genera with high relative abundances in source water which dropped almost to near zero in the subsequent sampling points. Coherent group **E** comprised particularly those genera enriched in aeration effluent but also in sedimentation effluent, presumably representing taxa with multiple niches. Members of cluster **D** were highly enriched in flocculation/sedimentation/pre-chlorination effluent, whereas those in cluster **C** peaked after sedimentation process. Cluster **F** and **A** represented groups of genera with peak abundance in sand filtration and post chlorination effluents, respectively.

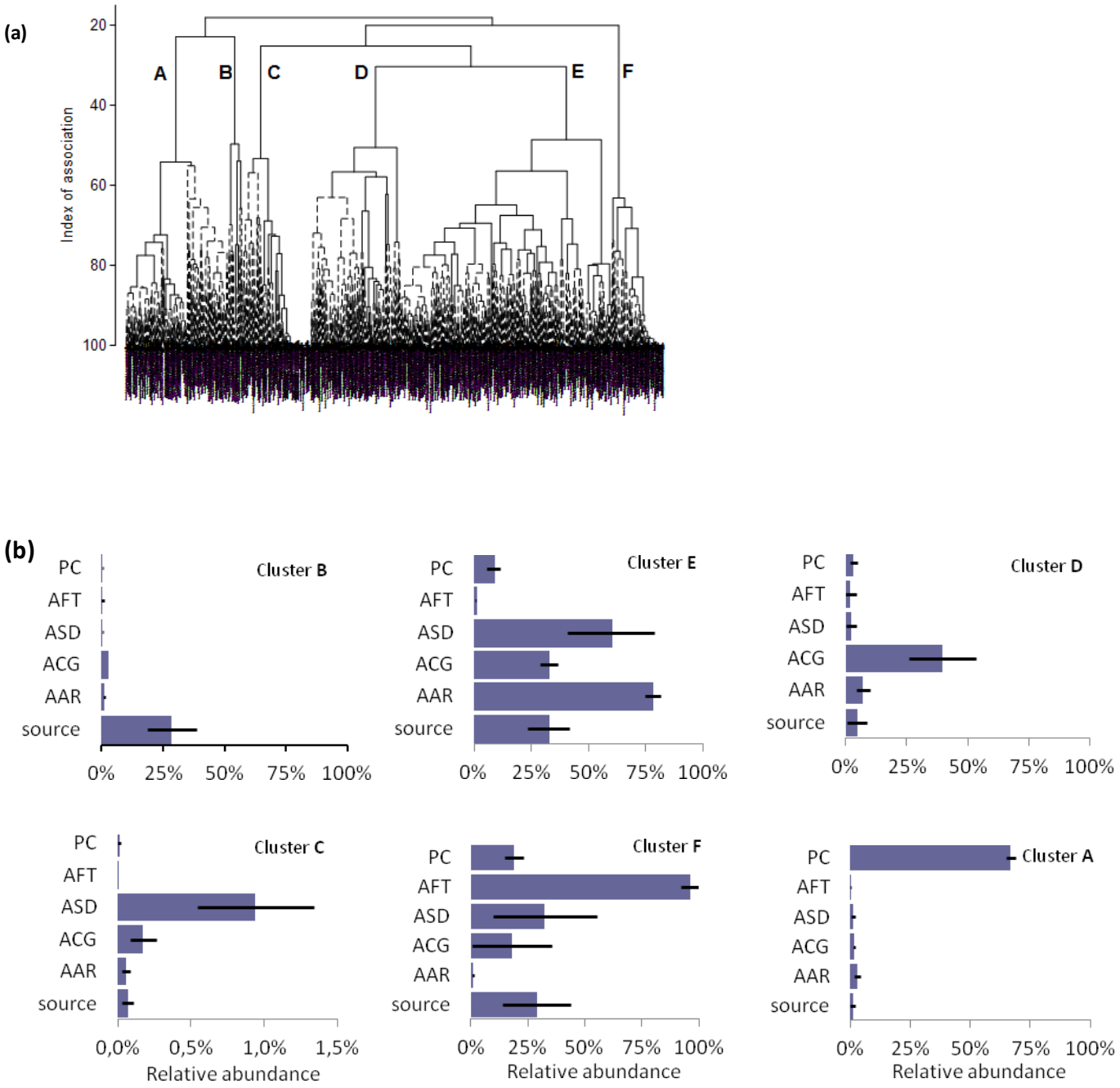


Figure 5: (**a**) A dendogram contains six major coherent groups of genera (A-F) from the pool of 1,003 genera that had significant coherent dynamics at *P*=0.02 across treatment barriers. Figure 5(**b**) shows the peak abundance of each coherent group (arranged according to the treatment barriers) corresponding to the effluent of treatment barrier in which it occurred. Abbreviation in the y-axis as in Figure 1.

4.3.5 Dominant genus predicts microbial diversity responses to drinking water treatment barriers.

The most dominant genus *Undibacterium* was highly dynamic and was present in high abundance (in some individual cases composed as much as 94% of the community) across all the samples. Data indicate that the population dynamics of *Undibacterium* in tropical DWTP responded directly to variations in the microbial diversity (Fig 6A), with a statististically significant correlation (Fig. 6B; Spearman correlation coefficient =-0.94, $p<0.005$), indicating that reduction of DWTP microbial diversity related to high abundance of *Undibacterium*, whereas high diversity hampered *Undibacterium* development. Other alpha diversity metrics were also applied and yielded comparable results for the given data set.

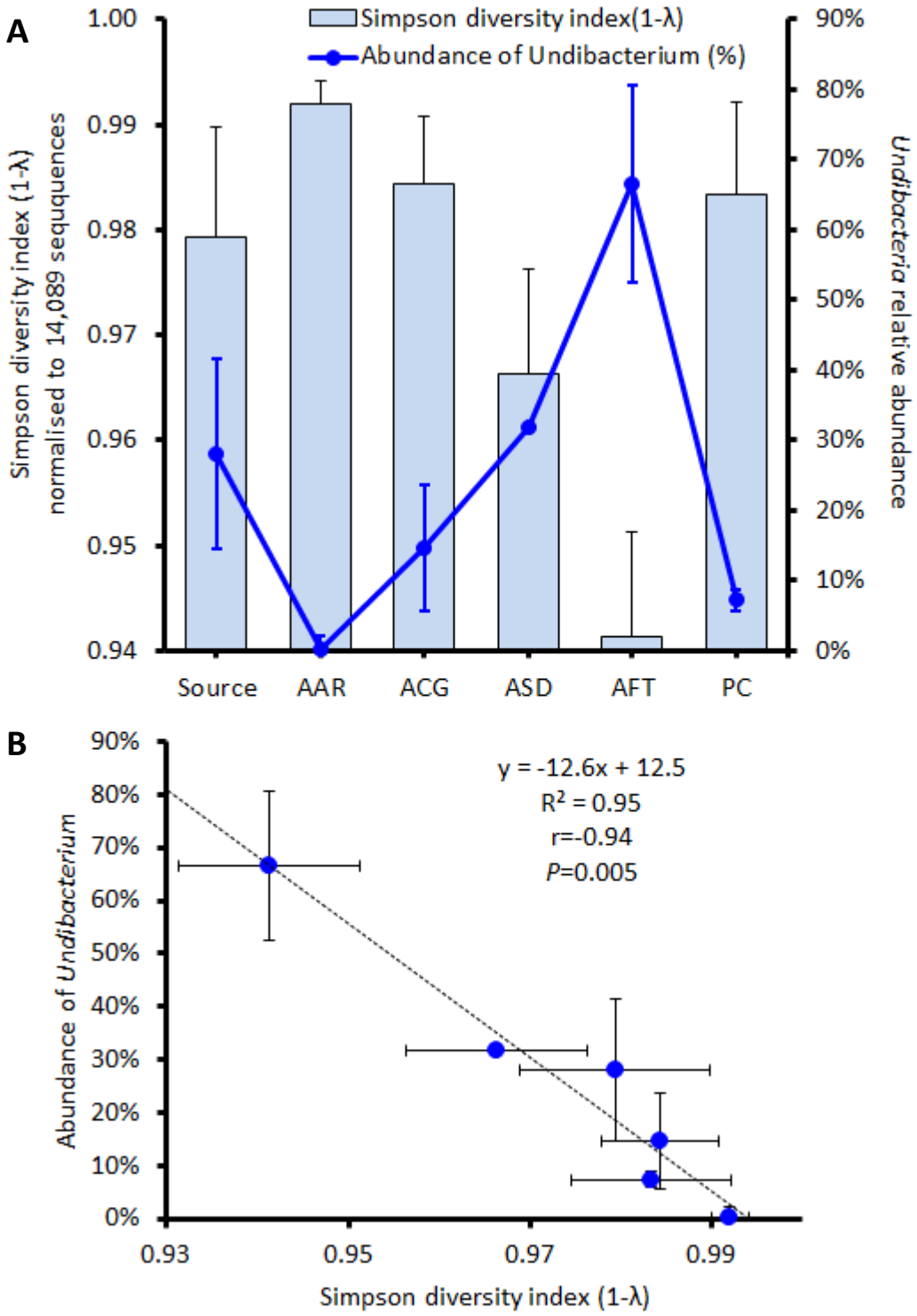


Figure 6. Fluctuations of mean alpha diversity and mean relative abundance (%) of genus *Undibacterium* (**A**), and their correlation (**B**) in tropical drinking water treatment plant. Abbreviations: AAR, aeration effluent; ACG, coagulation/flocculation/pre-chlorination effluent; ASD, sedimentation effluent; AFT, sand filtration effluent; PC, post chlorination effluent.

4.3.6 Increase in relative abundance of rare taxa post-chlorination

Thirty rare genera (<1% relative abundance) identified in the source water had a substantial increase in relative abundance in post-chlorination effluent (Table 1). Most of these genera were members of the phyla *Acidobacteria*, *Bacteroidetes* and *Proteobacteria*. The main representative of *Proteobacteria* subclasses were *Alphaproteobacteria*, *Betaproteobacteria*, *Delaproteobacteria* and *Gammaproteobacteria*. The genus *Methylophilus* had the highest percentage increase (294,900%), whereas genus *Legionella* had the lowest (51%).

Table 1: Phylogenetic analysis of rare taxa enriched in the post-chlorination effluent and their % change in mean (± SD) relative abundance between source water and post-chlorination effluent.

					% of total sequences (n=3)		
Phylum	**Class**	**Order**	**Family**	**Genus**	**Source**	**Post-chlorinat**	**% Change[c]**
Acidobacteria					**0.01 ± 0.0[b]**	**3.85 ± 0.03**	**38,400**
	Acidobacter	Subgroup 3	Unknown	*Candidatus Solibacter*	0.004 ± 0.0	2.01 ± 0.02	50,150
		Subgroup 3	Unknown	*Bryobacter*	0.002 ± 0.0	1.14 ± 0.004	56,900
	Holophagae	Holophagales	Holophagaceae	*Geothrix*	0.001 ± 0.0	0.70 ± 0.01	69,900
Bacteroidetes					**0.45 ± 0.01**	**21.74 ± 0.1**	**4,731**
	Cytophagia	Cytophagales	Cytophagaceae	Uncultured	0.16 ± 0.002	8.47 ± 0.1	5,193
		Cytophagales	Cytophagaceae	*Flectobacillus*	0.002 ± 0.0	3.11 ± 0.04	155,400
		Cytophagales	Cytophagaceae	*Emticicia*	0.001 ± 0.0	0.80 ± 0.01	79,900
	Sphingobacteriia	Sphingobacteriales	Chitinophagaceae	*Lacibacter*	0.09 ± 0.001	3.36 ± 0.03	3,633
		Sphingobacteriales	Chitinophagaceae	*Segetibacter*	0.001 ± 0.0	2.19 ± 0.02	218,900
		Sphingobacteriales	Chitinophagaceae	*Sediminibacterium*	0.03 ± 0.0	1.29 ± 0.01	4,200
		Sphingobacteriales	Chitinophagaceae	*Hydrotalea*	0.16 ± 0.003	0.68 ± 0.002	325
		Sphingobacteriales	Chitinophagaceae	*Flavisolibacter*	0.001 ± 0.0	0.59 ± 0.01	58,900
		Sphingobacteriales	Sphingobacteriaceae	*Mucilaginibacter*	0.001 ± 0.0	1.23 ± 0.004	122,900
Candidate division OD1[a]					**0.08 ± 0.001**	**2.76 ± 0.03**	**3,350**
Gemmatimonadetes					**0.02 ± 0.0**	**0.64 ± 0.01**	**3,100**
	Gemmatimonadetes	Gemmatimonadales	Gemmatimonadaceae	*Gemmatimonas*	0.02 ± 0.0	0.64 ± 0.01	3,100
Proteobacteria					**0.85 ± 0.01**	**18.61 ± 4.9**	**2,089**
	α-Proteobacteria	DB1-14	Unknown	Unknown	0.004 ± 0.0	0.78 ± 0.002	19,400
		Rhizobiales	Xanthobacteraceae	Uncultured	0.02 ± 0.0	0.52 ± 0.005	2,500
		Rhodobacterales	Rhodobacteraceae	*Rhodobacter*	0.03 ± 0.0	0.66 ± 0.002	2,100
		Rhodospirillales	Rhodospirillaceae	Uncultured	0.02 ± 0.0	0.62 ± 0.001	3,000
		Rickettsiales	SM2D12	Unknown	0.004 ± 0.0	2.24 ± 0.002	55,900
		Sphingomonadales	Sphingomonadaceae	*Novosphingobium*	0.04 ± 0.0	6.41 ± 0.03	15,925
		Sphingomonadales	Sphingomonadaceae	*Sphingomonas*	0.45 ± 0.01	2.68 ± 0.01	495
		Sphingomonadales	DSSF69	Unknown	0.14 ± 0.002	1.89 ± 0.005	1,250
	β-Proteobacteria	Burkholderiales	Comamonadaceae	Uncultured	0.44 ± 0.01	1.23 ± 0.003	179
		Methylophilales	Methylophilaceae	*Methylophilus*	0.002 ± 0.0	5.9 ± 0.1	294,900
		TRA3-20	Unknown	Unknown	0.04 ± 0.0	1.78 ± 0.02	4,350
	δ- Proteobacteria	Myxococcales	0319-6G20	Unknown	0.11 ± 0.001	1.31 ± 0.005	1,090
		Myxococcales	Cystobacteraceae	*Anaeromyxobacter*	0.01 ± 0.0	0.41 ± 0.003	4,000
	λ-Proteobacteria	Legionellales	Legionellaceae	*Legionella*	0.74 ± 0.01	1.12 ± 0.004	51
		Thiotrichales	Thiotrichaceae	Uncultured	0.001 ± 0.0	0.44 ± 0.01	43,900
SM2F11[a]					**0.09 ± 0.001**	**1.06 ± 0.01**	**1,077**
TM6[a]					**0.01 ± 0.0**	**0.87 ± 0.01**	**8,600**

[a]No classfication exists beyond Phylum level

[b]0.0 represents standard deviation with value < 10^{-3}

[c]Percentage change in relative abundance of genera between source and Post-chlorination samples calculated as:
% Change = ((Mean of Post-chlorination - Mean of Source water)/Mean of source water) × 100

4.4 Discussion

Analysis of microbial community dynamics in tropical drinking water treatment plant by sequencing the 16S rRNA gene at adequate depth is limited, despite high demand of microbiologically safe drinking water supply. Here, we screened microbial communities of a full

scale tropical drinking water treatment plant by sequencing the V4 and V5 hypervariable regions of 16S rRNA gene amplicon at appropriate sequencing depth (Fig. 1). High microbial diversity was revealed in tropical drinking water treatment plant as proven by genera detected from >26 phyla. However, 98-99% of this diverse microbial communities constituted low abundant genera (<1% relative abundance) and the majority had a prevalence <100% (supplementary Fig S5). Occurrence of low abundant genera in drinking water system can not be considered a methodological error because microbial communities were determined at the level of genus that is less likely to be sensitive to false taxon calls due to small sequencing errors (52). Additionally, low abundant genera were continuously detected across sample replicates, and in many discrete biological samples (53) considered in our investigation.

Although the dominating phyla in tropical drinking water treatment plant (Fig. 2) is consistent to the previous investigation (54), the composition of the top dominant phylum *Proteobacteria* was quite different from what have been reported so far for drinking water treatment plants. The phylum *Proteobacteria* was highly dominated by the class *Betaproteobacteria* in which the most dominant genus was *Undibacterium*. Regression anlysis showed that abundance of the genus *Undibacterium* could explain 73% of the variation of abundance of phylum *Proteobacteria* among the investigated samples, suggesting that the population of *Undibacterium* made up a stable and high percentage of the total *Proteobacteria* population in the investigated tropical drinking water treatment plant. Dominance of *Undibacterium* in DWTP has not been reported so far despite the vast number of studies performed on DWTP ((5,10–16,28,55–57), suggesting the importance of source water geographical location and the detection method applied. In this study, the genus *Undibacterium* demonstrated significant spatial fluctuation across the samples with the highest peak abundance (67% of total sequences/sample) immediately after significant bacterial taxa have been eliminated from DWTP by sand filtration (Fig. 5A). Increase of *Undibacterium* abundance immediately after sand filtration and their occurrence in environments with the lowest microbial diversity such as purified water (58) and finished drinking water (59), suggests that high microbial diversity limited their growth, and therefore, their appearance in water at elevated abundance may be interpreted as an indication of the presence of the lowest bacterial diversity (Fig.5B).

Source water microbial communities were significantly altered during the passage to different barriers along the drinking water treatment plant as shown by both alpha and beta diversity metrics (Table S2, Fig. 4 and 5). Microbial communities from source water to the sedimentation stage remained more or less the same even though pre-chlorination was integrated in the flocculation/coagulation stage, suggesting the importance of skipping a no impact treatment step from DWTP in order to cut operational costs on one hand and reduce disinfection byproducts on the other hand. Sand filtration was the most efficient step for removing substantial number of bacterial taxa from water intended for human consumption, as supported by the presence of only one dominant taxa and extremely low diversity compared to other treatment stages. This finding concurred with what has been reported elsewhere (10). Surprisingly, and in contrary to what was observed previously (10,13,50), microbial community composition after chlorination was quite different from that of sand filtration effluent and all samples before sand filtration, suggesting that microbial communities of this tropical DWTP were different from those detected in other climatic regions and mostly influenced by post-chlorination, and sand filtration barriers. It should be noted that the ability of alpha diversity metrics in detecting the influence of sand filtration and post chlorination in DWTP microbiome is of interest especially when one considers their insensitivity in detecting the presence of perturbation as reported elsewhere (60,61). In fact, a very severe impact posed by sand filtration and post-chlorination barriers on microbial communities of the tropical DWTP allowed even less sensitive parameters such as alpha diversity metrics to detect their effects.

Unexpectedly, genera considered rare (relative abundance below 1%) in source water and whose relative abundance were reduced after passing through the most efficient treatment barrier (sand filtration) discussed earlier in this study, showed a substantial enrichment in post-chlorination effluent (Table 1). *Candidatus Solibacter*, *Bryobacter*, *Geothrix, Emticicia* and *Sphingomonas* are known to have oligotrophic lifestyle (45,62–64), which seem to be favoured under resource limited environment (such as that of drinking water) because of high substrate affinities (45). On the other hand, *Geothrix*, *Rhodobacter*, *Anaeromyxobacteria* are iron respiring bacteria (65–68) and *Sediminibacterium* are iron oxidising bacteria (69). These bacterial taxa were enriched immediately after drinking water had entered into the iron/steel

pipes connecting the storage tank and the distribution networks. Iron of these type of pipes may have offered a suitable substrate for these taxa. In the same vein, bacterial genera capable of forming biofilm in drinking water pipes were enriched in post-chlorination effluent. These included *Sphingomonas, Methylophilus*, *Legionella*, *Rhodobacter*, *Sediminibacterium*, *Lacibacter* (9,46,70–74) among others (Table 1). It can be hypothesized that their increase in abundance in post-chlorination effluent may be due to detachment of biofilm from the surface of the drinking water pipes into the bulk water. Additionally, increase in relative abundance of taxa belonging to *Betaproteobacteria* in post-chlorination may be associated with chlorine addition (75). It should be noted that amplification of rare abundant genera substatially changed the microbial composition in post-chlorination effluent resulting to a shift in microbial community structure observed previously (Fig. 4), and explain the existence of naturally occuring rare taxa in drinking water designed for human consumption.

Rare taxa enrichment in post-chlorination effluent is evidence supporting the previous idea that rare taxa detected in this study are not artifacts caused by sequencing errors rather than naturally occurring taxa in the tropical DWTP. This enrichment provide an insight into existence of completely different niche in post chlorination effluent as compared to that of source water, and support the idea that rare taxa are ecologically active (76). A major ecological question is whether enrichment of rare taxa is due to contamination by leaking pipes and/or regrowth within the drinking water system immediately after chlorination. For a contamination, one would expect significant increase in relative abundance for dominant taxa as well. However, this was not the case in this investigation as many of the most dominating taxa were reduced after chlorination (see Fig. 3). Furthermore, taxa known to occur exclusively in oligotrophic environment such as *Emtiticia* (64) and phylum *Acidobacteria* (45,77) were among the taxa enriched tremendously after chlorination, suggesting that no signature of leaking pipes that would lead to drinking water contamination. On the other hand most members of *Proteobacteria* including *Betaproteobacteria* and *Gammaproteobacteria* ((46,75,78) that are known to be favoured by high free residual chlorine concentration were among the taxa increased in post-chlorination effluent (Table 1). This increase can be assumed as a response to the added chlorine, and it raised doubt about the microbiological safety of the finished water

because members of these taxa can cause threat to public health. Since taxa such as *Sphingomonas* are known for biofilm formation in the drinking water pipes (46), a contribution of biofilm to the observed increase in abundance of rare taxa in post-chlorination effluent can not be excluded.

One of the primary goals of monitoring complex microbial communities of DWTP using 16S rRNA gene amplicon deep sequencing is to ensure that water supplied to end users is free from microbial pathogens. With this concept in mind, apt sequencing depth was applied to ensure that all bacterial taxa present in the DWTP were detected (Fig.1A and B). However, it should be mentioned that accurate pathogen detection in drinking water may not be possible due to their low abundance and primer biases. Nevertheless, the universal primers used in this study allowed detection of several genera known to encompass pathogenic species – including *Legionella, Sphingomonas, Mycobacterium* and *Rickettsia* – in the finished water at low abundance, suggesting the use of primers more specific to pathogenic species of these genera in order to confirm their existence and detail their species composition. In the same vein, taxa known to be reservoirs of antibiotic resistance – such as *Bosea* and *Afipia* (49), *Sphingomonadaceae* (48), *Acinetobacter* (47) and *Pseudomonas* (50) – were also detected in the finished water, suggesting that finished water could be an important vector for antibiotic resistant bacteria.

Variability of genera across the sampling points was extremely high with six significantly distinct patterns that revealed the sequential progression of genera across treatment barriers of the DWTP (Fig 5). This complex dynamics patterns were independent of taxonomic relatedness as taxonomically distinct genera had similar patterns across the sampling points. Each dynamic pattern (contained distinct genera composition) observed in this study had peak abundance in at least one of the treatment barrier effluents, suggesting the presence of a secondary succession in the DWTP. The secondary succession in the DWTP is an indication that each treatment barrier modifies to some extent the water quality to allow other genera that are better suited to that modification to succeed. Microbial succession in drinking water may be a public health concern. However, succession patterns observed here allow prediction of changes

of specific genera especially those related with pathogenicity for better management of microbiological water quality in DWTP.

In summary, this study identified sand filtration and post chlorination barriers as the main treatment steps responsible for the microbial diversity and community structure alteration in tropical DWTP. While the former efficiently eliminated significant numbers of microbial genera from the water intended for drinking, the latter facilitated the enrichment of taxa including those that are a threat to public health. The novel correlation of the relative abundance of the genus *Undibacterium* with microbial diversity may indicate an operational possibility to evaluate efficiently the microbiological safety and quality of the finished water in tropical DWTP. Microbial successions as revealed by the coherent dynamics of taxa across the treatment barriers may in future allow the prediction of changes of specific taxa across treatment barriers of relevance especially for those known to be a threat to public health.

4.5 Literature cited

1. Hellard ME, Sinclair MI, Forbes AB, Fairley CK. A randomized, blinded, controlled trial investigating the gastrointestinal health effects of drinking water quality. Environ Health Perspect. 2001;109(8):773–8.
2. Kristemann T, Claßen T, Koch C, Dangendorf F, Gebel J, Vacata V, et al. Microbial load of drinking water reservoir tributaries during extreme rainfall and runoff microbial load of drinking water reservoir tributaries during extreme rainfall and runoff. Appl Environ Microbiol. 2002;68(5):2188–97.
3. Hua W, Bennett ER, Letcher RJ. Ozone treatment and the depletion of detectable pharmaceuticals and atrazine herbicide in drinking water sourced from the upper Detroit River, Ontario, Canada. Water Res. 2006;40(12):2259–66.
4. Kosaka K, Asami M, Matsuoka Y, Kamoshita M, Kunikane S. Occurrence of perchlorate in drinking water sources of metropolitan area in Japan. Water Res. 2007;41(15):3474–82.
5. Lin W, Yu Z, Zhang H, Thompson IP. Diversity and dynamics of microbial communities at each step of treatment plant for potable water generation. Water Res. 2014;52:218–30.
6. Reweta WSJ, Sampath RK. Performance evaluation of urban water supply in Tanzania: the case of Dar Es Salaam City. Int J Water Resour Dev. 2000;16(3):407–21.
7. Marobhe NJ. Critical review of water supply services in urban and rural areas of Tanzania. Water Policy. 2008;10(1):57–71.
8. Komakech HC, Van der Zaag P, Van Koppen B. The last will be first: Water transfers from agriculture to cities in the Pangani River Basin, Tanzania. Water Altern. 2012;5(3):700–20.
9. Wang H, Hu C, Hu X, Yang M, Qu J. Effects of disinfectant and biofilm on the corrosion of cast iron pipes in a reclaimed water distribution system. Water Res. 2012;46(4):1070–8.

10. Pinto AJ, Xi C, Raskin L. Bacterial community structure in the drinking water microbiome is governed by filtration processes. Environ Sci Technol. 2012;46(16):8851–9.
11. Lautenschlager K, Hwang C, Ling F, Liu WT, Boon N, Köster O, et al. Abundance and composition of indigenous bacterial communities in a multi-step biofiltration-based drinking water treatment plant. Water Res. 2014;62:40–52.
12. Vaz-Moreira I, Egas C, Nunes OC, Manaia CM. Bacterial diversity from the source to the tap: a comparative study based on 16S rRNA gene-DGGE and culture-dependent methods. FEMS Microbiol Ecol. 2013;83(2):361–74.
13. Eichler S, Christen R, Höltje C, Bötel J, Brettar I, Mehling A, et al. Composition and dynamics of bacterial communities of a drinking water supply system as assessed by RNA- and DNA-Based 16S rRNA gene fingerprinting. Appl Environ Microbiol. 2006;72(3):1858–72.
14. Zeng D-N, Fan Z-Y, Chi L, Wang X, Qu W-D, Quan Z-X. Analysis of the bacterial communities associated with different drinking water treatment processes. World J Microbiol Biotechnol. 2013;29(9):1573–84.
15. Liao X, Chen C, Wang Z, Chang C-H, Zhang X, Xie S. Bacterial community change through drinking water treatment processes. Int J Environ Sci Technol. 2015;12(6):1867–74.
16. Norton CD, Lechevallier MW. A pilot study of bacteriological population changes through potable water treatment and distribution. Appl Environ Microbiol. 2000;66(1):268–76.
17. Huang K, Zhang X-X, Shi P, Wu B, Ren H. A comprehensive insight into bacterial virulence in drinking water using 454 pyrosequencing and Illumina high-throughput sequencing. Ecotoxicol Environ Saf. 2014;109:15–21.
18. Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Lozupone CA, Turnbaugh PJ, et al. Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. Proc Natl Acad Sci USA. 2011;108:4516–22.
19. Kumaraswamy R, Amha YM, Anwar MZ, Henschel A, Rodr J, Ahmad F. Molecular analysis for screening human bacterial pathogens in municipal wastewater treatment and reuse. Environ Sci Technol. 2014;48:11610-19.
20. Cai L, Zhang T. Detecting human bacterial pathogens in wastewater treatment plants by a high-throughput shotgun sequencing technique. Environ Sci Technol. 2013;47:5433–41.
21. Newton RJ, Mclellan SL, Dila DK, Vineis JH, Morrison HG, Eren AM, et al. Sewage Reflects the Microbiomes of Human Populations. MBio. 2015;6(2):1–9.
22. Hong C, Si Y, Xing Y, Li Y. Illumina MiSeq sequencing investigation on the contrasting soil bacterial community structures in different iron mining areas. Environ Sci Pollut Res. 2015;
23. Yang C, Li Y, Zhou B, Zhou Y, Zheng W, Tian Y, et al. Illumina sequencing-based analysis of free-living bacterial community dynamics during an Akashiwo sanguine bloom in Xiamen sea, China. Sci Rep. 2015;5:8476.
24. Staley C, Unno T, Gould TJ, Jarvis B, Phillips J, Cotner JB, et al. Application of Illumina next-generation sequencing to characterize the bacterial community of the Upper Mississippi River. J Appl Microbiol. 2013;115(5):1147–58.
25. Staley C, Gould TJ, Wang P, Phillips J, Cotner JB, Sadowsky MJ. Bacterial community structure is indicative of specific types of contamination in the Upper Mississippi River. Front. Microbiol. 5:524.

26. Hamady M, Knight R, Stern A, Mick E, Tirosh I. Microbial community profiling for human microbiome projects : Tools , techniques , and challenges. Genome Res. 2009;19:1141–52.
27. Nkwengulila G, Kigadye E. Occurrence of digenean larvae in freshwater snails in the Ruvu basin, Tanzania. Tanzania J Sci. 2009;31(2).
28. Albers CN, Ellegaard-Jensen L, Harder CB, Rosendahl S, Knudsen BE, Ekelund F, et al. Groundwater Chemistry Determines the Prokaryotic Community Structure of Waterworks Sand Filters. Environ Sci Technol. 2015;49(2):839–46.
29. Barbeau K, Kujawinski EB, Moffett JW. Remineralization and recycling of iron, thorium and organic carbon by heterotrophic marine protists in culture. Aquat Microb Ecol. 2001;24(1):69–81.
30. Camarinha-Silva A, Jáuregui R, Chaves-Moreno D, Oxley APA, Schaumburg F, Becker K, et al. Comparing the anterior nare bacterial community of two discrete human populations using Illumina amplicon sequencing. Environ Microbiol. 2014;16:2939–52.
31. Schwieger F, Tebbe CC. A new approach to utilize PCR – single-strand-conformation polymorphism for 16S rRNA gene-based microbial community analysis. Appl Environ Microbiol. 1998;64(12):4870–6.
32. Hamady M, Walker JJ, Harris JK, Gold NJ, Knight R. Error-correcting barcoded primers for pyrosequencing hundreds of samples in multiplex. Nat Methods. 2008;5(3):235–7.
33. Kittelmann S, Seedorf H, Walters W a., Clemente JC, Knight R, Gordon JI, et al. Simultaneous Amplicon Sequencing to Explore Co-Occurrence Patterns of Bacterial, Archaeal and Eukaryotic Microorganisms in Rumen Microbial Communities. PLoS One. 2013;8(2).e47879.
34. Caporaso JG, Lauber CL, Walters W a, Berg-Lyons D, Huntley J, Fierer N, et al. Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. ISME J. 2012;6(8):1621–4.
35. Van Dijk EL, Jaszczyszyn Y, Thermes C. Library preparation methods for next-generation sequencing: Tone down the bias. Exp Cell Res. 2014;322(1):12–20.
36. Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, et al. Introducing mothur: Open-source, platform-independent, community-supported software for describing and comparing microbial communities. Appl Environ Microbiol. 2009;75(23):7537–41.
37. Edgar RC, Haas BJ, Clemente JC, Quince C, Knight R. UCHIME improves sensitivity and speed of chimera detection. Bioinformatics. 2011;27(16):2194–200.
38. Pruesse E, Peplies J, Glöckner FO. SINA: Accurate high-throughput multiple sequence alignment of ribosomal RNA genes. Bioinformatics. 2012;28(14):1823–9.
39. Cole JR, Wang Q, Cardenas E, Fish J, Chai B, Farris RJ, et al. The Ribosomal Database Project: Improved alignments and new tools for rRNA analysis. Nucleic Acids Res. 2009;37(SUPPL. 1):141–5.
40. Robertson CE, Harris JK, Wagner BD, Granger D, Browne K, Tatem B, et al. Explicet: Graphical user interface software for metadata-driven management, analysis and visualization of microbiome data. Bioinformatics. 2013;29(23):3100–1.

41. Somerfield PJ, Burton M, Sanderson WG. Analyses of sublittoral macrobenthic community change in a marine nature reserve using similarity profiles (SIMPROF). Mar Environ Res. 2014;51–8.
42. Somerfield PJ, Clarke KR. Inverse analysis in non-parametric multivariate analyses: Distinguishing groups of associated species which covary coherently across samples. J Exp Mar Bio Ecol. 2013;449:261–73.
43. Parks DH, Beiko RG. Identifying biologically relevant differences between metagenomic communities. Bioinformatics. 2010;26(6):715–21.
44. Saker ML, Nogueira ICG, Vasconcelos VM, Neilan B a., Eaglesham GK, Pereira P. First report and toxicological assessment of the cyanobacterium *Cylindrospermopsis raciborskii* from Portuguese freshwaters. Ecotoxicol Environ Saf. 2003;55(2):243–50.
45. Fierer N, Bradford M a., Jackson RB. Toward an ecological classification of soil bacteria. Ecology. 2007;88(6):1354–64.
46. Koskinen R, TAli-Vehmas, Kämpfer P, Laurikkala M, Tsitko I, Kostyal E, et al. Characterization of *Sphingomonas* isolates from Finnish and Swedish drinking water distribution systems. J Appl Microbiol. 2000;89(4):687–96.
47. Gales a C, Jones RN, Forward KR, Liñares J, Sader HS, Verhoef J. Emerging importance of multidrug-resistant *Acinetobacter* species and *Stenotrophomonas maltophilia* as pathogens in seriously ill patients: geographic patterns, epidemiological features, and trends in the SENTRY Antimicrobial Surveillance Program (1997-1999). Clin Infect Dis. 2001;32 (Suppl 2):S104–13.
48. Vaz-Moreira I, Nunes OC, Manaia CM. Diversity and antibiotic resistance patterns of *Sphingomonadaceae* isolates from drinking water. Appl Environ Microbiol. 2011;77(16):5697–706.
49. Falcone-Dias MF, Vaz-Moreira I, Manaia CM. Bottled mineral water as a potential source of antibiotic resistant bacteria. Water Res. 2012;46(11):3612–22.
50. Vaz-Moreira I, Nunes OC, Manaia CM. Diversity and antibiotic resistance in *Pseudomonas spp*. from drinking water. Sci Total Environ. 2012;426:366–74.
51. Hunter PD. Cyanobacterial toxins and human health. Symp Ser Soc Appl Microbiol. 1998;27:35S – 40S.
52. Zhou Y, Gao H, Mihindukulasuriya KA, La Rosa PS, Wylie KM, Vishnivetskaya T, et al. Biogeography of the ecosystems of the healthy human body. Genome Biol. 2013;14(1):R1.
53. Reeder J, Knight R. The 'rare biosphere': a reality check. Nat Methods. 2009;6(9):636–7.
54. Revetta RP, Pemberton A, Lamendella R, Iker B, Santo Domingo JW. Identification of bacterial populations in drinking water using 16S rRNA-based sequence analyses. Water Res. 2010;44(5):1353–60.
55. White CP, DeBry RW, Lytle DA. Microbial survey of a full-scale, biologically active filter for treatment of drinking water. Appl Environ Microbiol. 2012;78(17):6390–4.
56. Bai Y, Liu R, Liang J, Qu J. Integrated metagenomic and physiochemical analyses to evaluate the potential role of microbes in the sand filter of a drinking water treatment system. PLoS One. 2013;8(4).e61011

57. Hoefel D, Monis PT, Grooby WL, Andrews S, Saint CP. Profiling bacterial survival through a water treatment process and subsequent distribution system. J Appl Microbiol. 2005;99(1):175–86.
58. Eder W, Wanner G, Ludwig W, Busse HJ, Ziemke-Kägeler F, Lang E. Description of *Undibacterium oligocarboniphilum* sp. nov., isolated from purified water, and *Undibacterium pigrum* strain CCUG 49012 as the type strain of *Undibacterium parvum* sp. nov., and emended descriptions of the genus *Undibacterium* and the species *Undibacterium pigrum*. Int J Syst Evol Microbiol. 2011;61(2):384–91.
59. Kämpfer P, Rosselló-Mora R, Hermansson M, Persson F, Huber B, Falsen E, et al. *Undibacterium pigrum* gen. nov., sp. nov., isolated from drinking water. Int J Syst Evol Microbiol. 2007;57(7):1510–5.
60. Hartmann M, Widmer F. Community structure analyses are more sensitive to differences in soil bacterial communities than anonymous diversity indices. Appl Environ Microbiol. 2006;72(12):7804–12.
61. McCoy CO, Matsen FA. Abundance-weighted phylogenetic diversity measures distinguish microbial community states and are robust to sampling depth. PeerJ. 2013;1:e157.
62. Naether A, Foesel BU, Naegele V, Wüst PK, Weinert J, Bonkowski M, et al. Environmental factors affect acidobacterial communities below the subgroup level in grassland and forest soils. Appl Environ Microbiol. 2012;78(20):7398–406.
63. Ohta H. Kinetic analysis of ferulic acid degradation by oligotrophic *Sphingomonas* sp. S213 during Growth in Batch and Continuous Cultures. Microbes Environ. 2001;16(1):9–17.
64. Saha P, Chakrabarti T. *Emticicia oligotrophica* gen. nov., sp. nov., a new member of the family '*Flexibacteraceae*', phylum *Bacteroidetes*. Int J Syst Evol Microbiol. 2006;56(Pt 5):991–5.
65. Coates JD, Ellis DJ, Gaw C V, Lovley DR. *Geothrix ferrnentans*. Int J Syst Bacteriol. 1999;49:1615–22.
66. He Q, Sanford R a. Characterization of Fe (III) Reduction by Chlororespiring *Anaeromxyobacter dehalogenans* Characterization of Fe (III) Reduction by Chlororespiring *Anaeromxyobacter dehalogenans*. 2003;69(5):2712.
67. Dobbin P, Warren LH, Cook NJ, Mcewan AG, Powell AK, Richardson DJ. Dissimilatory iron(lll) reduction by R*hodobacter capsulatus*. Microbiol. 1996;142:765–74.
68. Treude N, Rosencrantz D, Liesack W, Schnell S. Strain FAc12, a dissimilatory iron-reducing member of the *Anaeromyxobacter* subgroup of *Myxococcales*. FEMS Microbiol Ecol. 2003;44(2):261–9.
69. Wang J, Sickinger M, Ciobota V, Herrmann M, Rasch H, Rösch P, et al. Revealing the microbial community structure of clogging materials in dewatering wells differing in physico-chemical parameters in an open-cast mining area. Water Res. 2014;63:222–33.
70. Buse HY, Lu J, Lu X, Mou X, Ashbolt NJ. Microbial diversities (16S and 18S rRNA gene pyrosequencing) and environmental pathogens within drinking water biofilms grown on the common premise plumbing materials unplasticized polyvinylchloride and copper. FEMS Microbiol Ecol. 2014;88(2):280–95.
71. Declerck P. Biofilms: The environmental playground of *Legionella pneumophila*. Environ Microbiol. 2010;12(3):557–66.

72. Venugopalan VP, Kuehn M, Hausner M, Springael D, Wilderer PA., Wuertz S. Architecture of a Nascent *Sphingomonas sp*. biofilm under varied hydrodynamic conditions. Appl Environ Microbiol. 2005;71(5):2677–86.
73. Wilkinson DA., Chacko SJ, Vénien-Bryan C, Wadhams GH, Armitage JP. Regulation of flagellum number by FliA and FlgM and role in biofilm formation by Rhodobacter sphaeroides. J Bacteriol. 2011;193(15):4010–4.
74. Van Der Kooij D, Veenendaal HR, Scheffer WJH. Biofilm formation and multiplication of Legionella in a model warm water system with pipes of copper, stainless steel and cross-linked polyethylene. Water Res. 2005;39(13):2789–98.
75. Mathieu L, Bouteleux C, Fass S, Angel E, Block JC. Reversible shift in the α-,β- and λ-*proteobacteria* populations of drinking water biofilms during discontinuous chlorination. Water Res. 2009;43(14):3375–86.
76. Wilhelm L, Besemer K, Fasching C, Urich T, Singer GA., Quince C, et al. Rare but active taxa contribute to community dynamics of benthic biofilms in glacier-fed streams. Environ Microbiol. 2014;16(8):2514–24.
77. Sims A, Zhang Y, Gajaraj S, Brown PB, Hu Z. Toward the development of microbial indicators for wetland assessment. Water Res. 2013;47(5):1711–25.
78. Hwang C, Ling F, Andersen GL, LeChevallier MW, Liu WT. Microbial community dynamics of an urban drinking water distribution system subjected to phases of chloramination and chlorination treatments. Appl Environ Microbiol. 2012;78(22):7856–65.

4.6 Supplementary materials

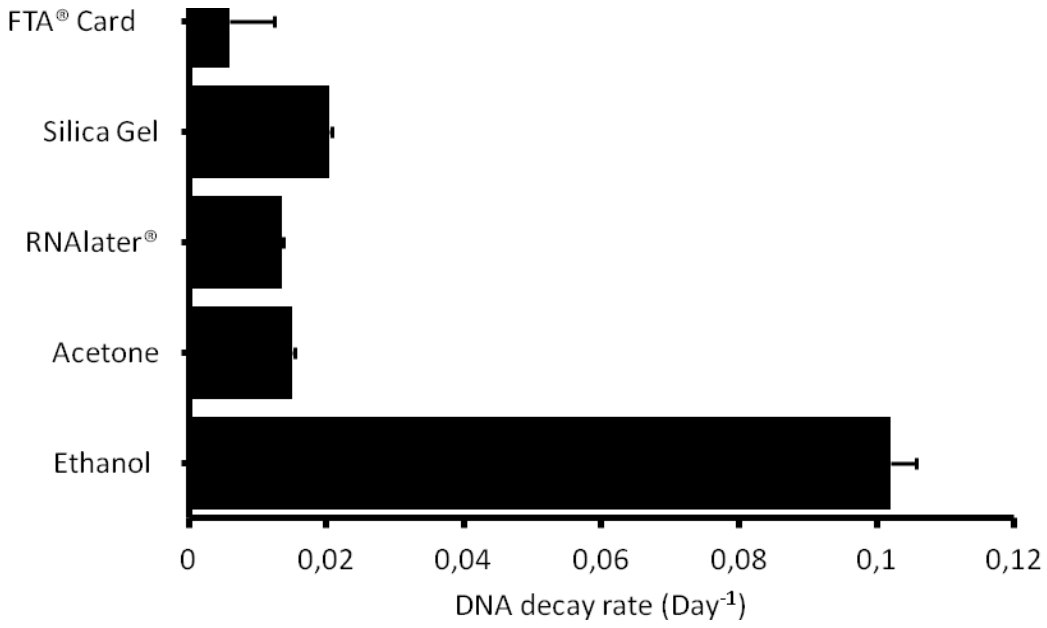


Figure S1: Comparisons of non-freezing DNA storage methods with respect to DNA decay over time (>8 weeks) at room temperature. Compared methods were previously applied in various studies including Rissanen *et al.*, (2010) (Ethanol and RNAlater®), Fukatsu (1999) (Acetone), Bainard *et al.*, (2010) (Silica gel) and Saieg *et al.* (2012) (FTA Cards). Of all the methods compared, DNA on FTA card had the lowest decay rate averaging to 0.006 per day.

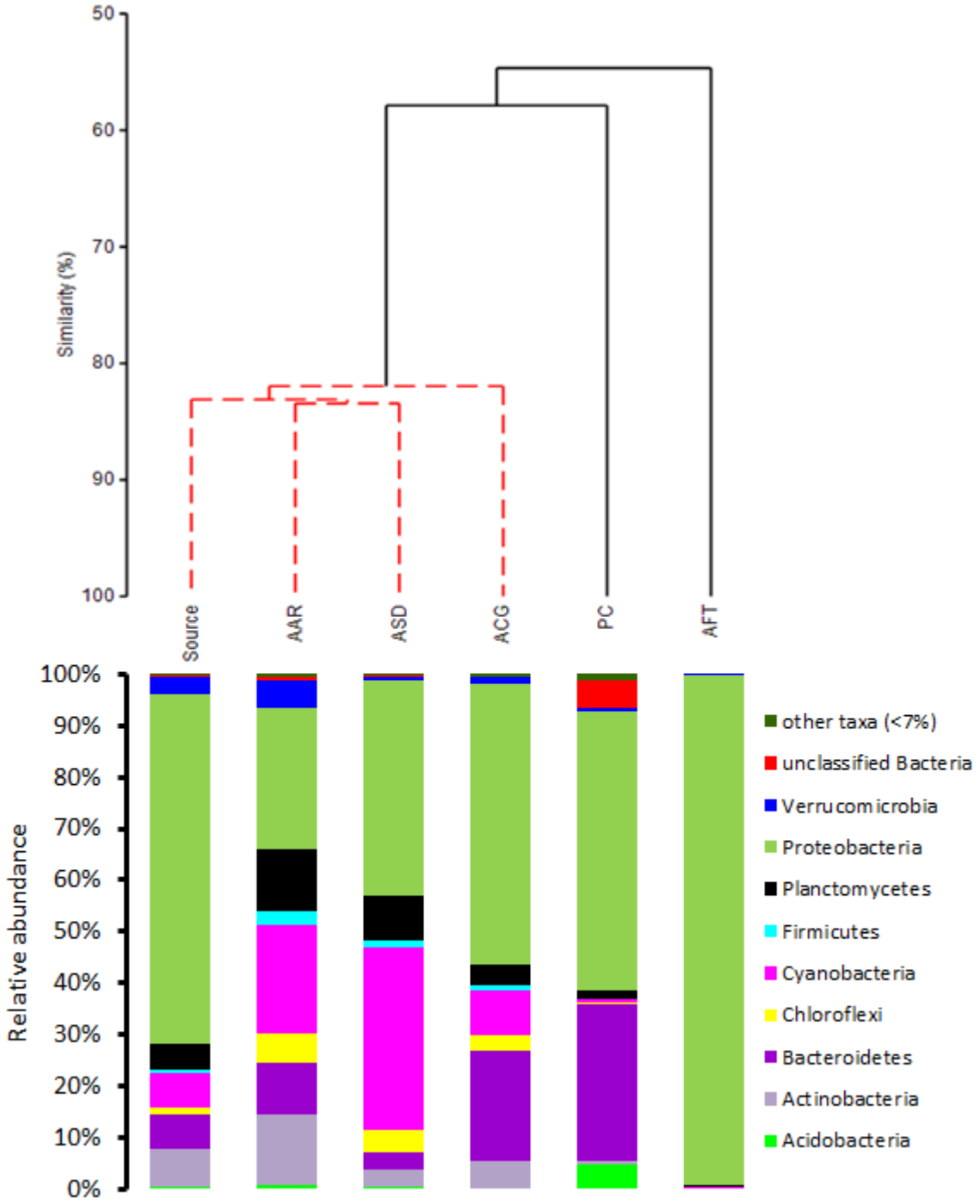


Figure S2: Microbial community structure and composition in tropical drinking water treatment plant. Community structure indicated that source water, aeration (AAR), coagulation/flocculation/pre-chlorination (ACG), and sedimentation (ASD) had similar microbial composition, indicating that source water seeds bacteria to drinking water treatment plant. However, Sand filtration (AFT) and Post-chlorination had substantial discrete microbial composition between each other, and the rest of the samples. It should be emphasized that average values from triplicate samples were used to construct the displayed graph.

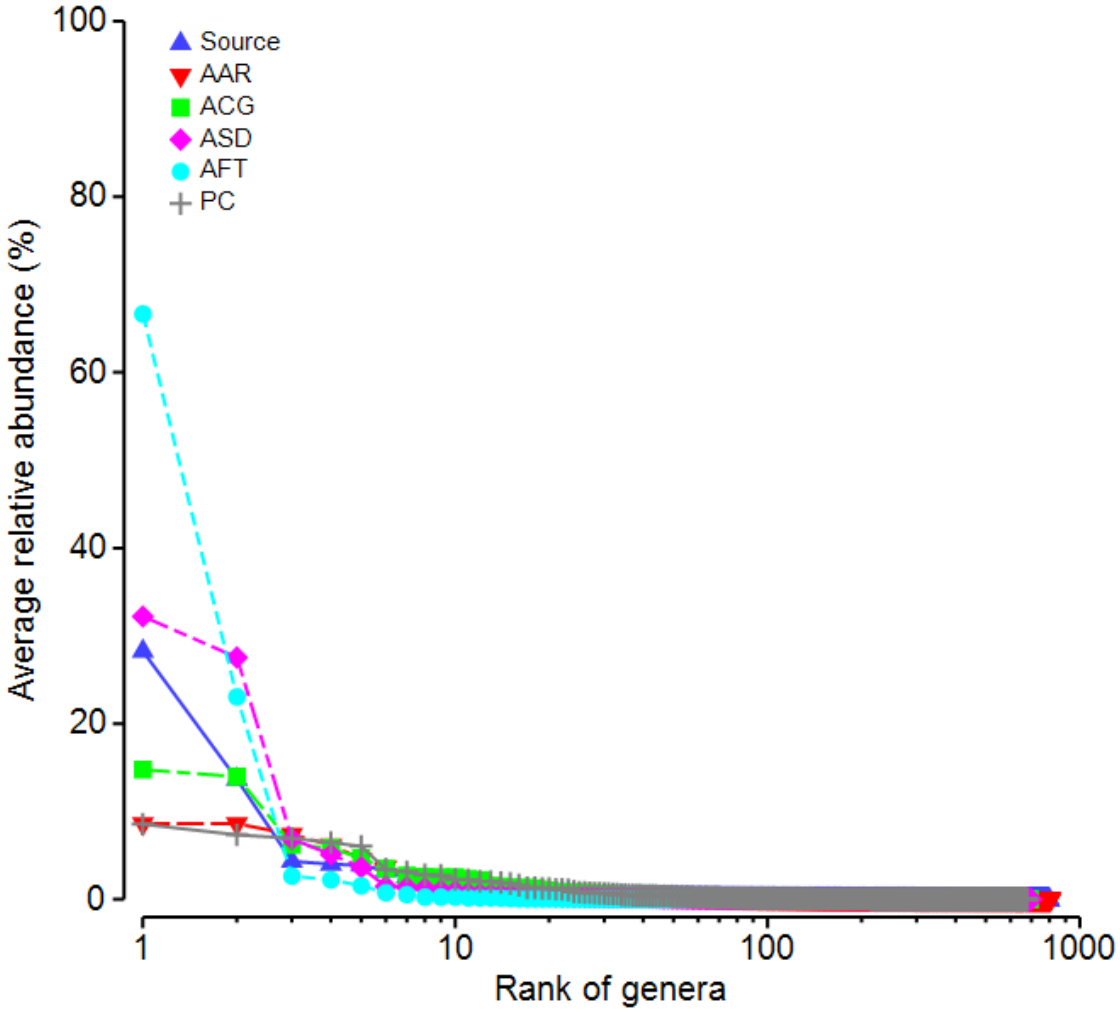


Figure S3: Rank abundance curves showing the distribution of genera in the investigated samples relative to the average abundance. x-axis shows the ranking of genera starting from those with the highest relative abundance to the lowest whereas, y-axis shows the average relative abundance of each genera. The smallest number of genera were dominating the samples while majority were low abundant (<1 relative abundance) taxa. For abbreviation see Fig. S3.

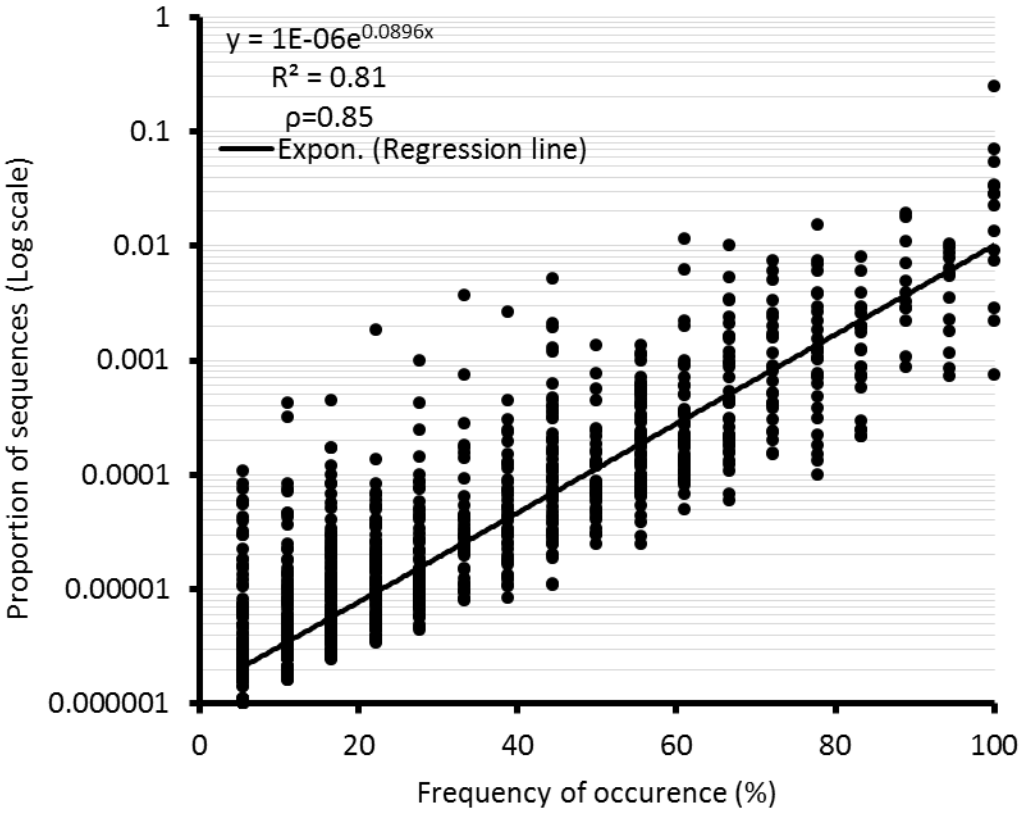


Figure S4: Relationship between prevalence and proportion of sequences for the detected genera in tropical drinking water treatment plant. Dominating genera (relative abundance above 0.01) tended to have the highest prevalence while majority of the low abundant genera tended to occur in more than one sample and some with cosmopolitan life style.

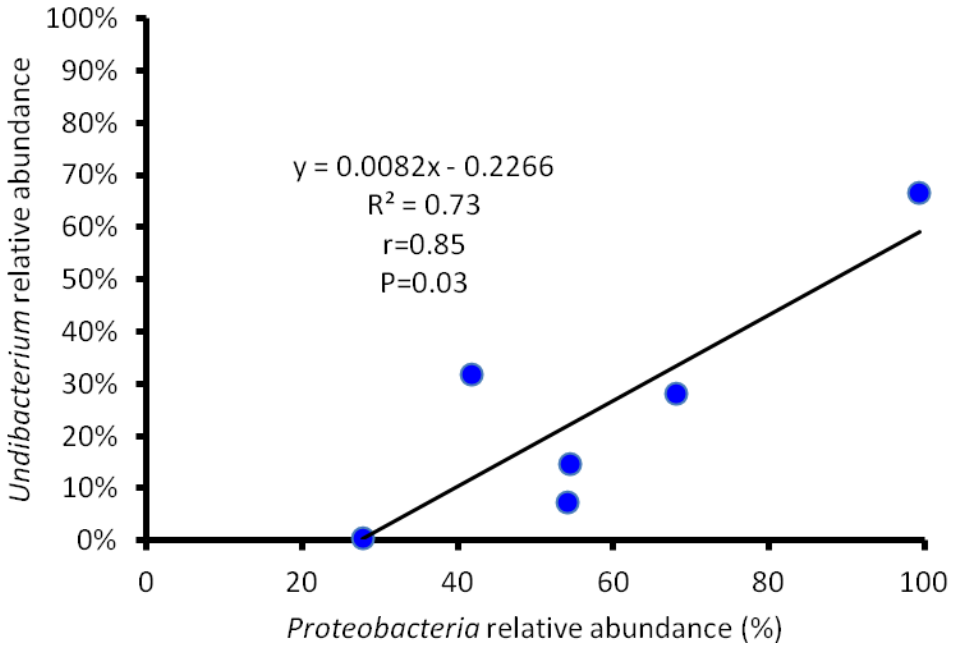


Figure S5: Relationship between average abundances of *Undibacterium* and *Proteobacteria*. About 73% of the variation of *Proteobacteria* abundance was explained by the abundance of *Undibacterium*. The Pearson correlation coefficient (r=0.85) and probability value (P=0.03) of the observed relationship is shown.

Table S1. Efficiency of bacterial cells recovery from polycarbonate membrane filters.

Filter #	Cell counts in the original sample /ml	Cell counts retrieved from polycarbonate membrane /ml	Recovery Efficiency %
1	1.90E+06	1.90E+06	99.89
2	1.87E+06	1.87E+06	100.00
3	1.50E+06	1.49E+06	99.33
4	1.81E+06	1.80E+06	99.38
5	1.63E+07	1.61E+07	98.77
6	3.10E+06	3.09E+06	99.68
7	2.70E+07	2.69E+07	99.63
8	2.10E+06	2.09E+06	99.38
9	2.00E+07	2.00E+07	99.90
10	1.90E+07	1.89E+07	99.36

Table S2: 16S rRNA sequences, alpha diversity metrics, and total phyla data set for the surveyed sampling points.

Site	Mean (range) for:[a] Sequences	Observed OTUs	Chao1	ACE	Shannon (H')	Simpson (1/λ)	Good's coverage	Phyla
Source	63,458	1,860	2,470	2,586	8.5	83	0.95	24
	(60,727-68,767)	(1,175-2,362)	(1,503-3,245)	(1,528-3,474)	(7.3-9.0)	(24-157)	(0.93-0.97)	
AAR	46,845	1,796	2,370	2,514	8.8	145	0.96	26
	(18,256-68,244)	(819-2,389)	(822-3,230)	(829-3,478)	(7.9-8.8)	(82-204)	(0.93-1.0)	
ACG	39,644	1,378	1,689	1,746	8.3	90	0.97	22
	(15,427-52,953)	(712-1,863)	(712-2,423)	(712-2,542)	(7.3-8.9)	(36-157)	(0.95-1.0)	
ASD	43,764	1,505	1,889	1,984	8.1	86	0.97	21
	(17,142-59,532)	(835-2,713)	(836-3,715)	(840-4,005)	(6.7-9.5)	(12-154)	(0.92-1.0)	
AFT	49,163	961	1,131	1,144	7.1	21	0.98	13
	(39,050-60,105)	(780-1,131)	(892-1,376)	(911-1,385)	(6.3-7.5)	(9-31)	(0.98-0.99)	
PC	62,774	2047	2,816	2,970	8.6	104	0.94	21
	(55,544-67,193)	(1,453-2,530)	(1,922-3,539)	(2,031-3,732)	(7.5-9.4)	(29-195)	(0.93-0.96)	

[a]Alpha diversity data were obtained after normalizing OTUs to 14,083 sequences.

Appendix

Appendix 1A: Physico-chemical and microbiological data set for dry season, nd = not determined; BA = bacterial abundance; TC = total coliforms, HPC = heterotrophic plate counts, Temp = temperature, cond = conductivity, DO = dissolved oxygen, BOD = five day's biological oxygen demand.

Site Id	Sampling dates	Temp °C	pH	Cond µS	DO mg/l	TSS mg/l	BOD mg/L	NH_4^+ mg/l	PO_4^{3-} mg/l	NO_3^- mg/l	NO2 mg/l	Hardness mg/l	Chloride mg/l	HPC cfu/ml	TC cfu/ml	BA cells/ml
Pr1D	8/10/2012	19.2	7.93	64.9	9.9	0.02	1.8	0.30	10	2	0.02	53	9	4160	**nd**	1.02E+06
	15/10/2012	19.2	8.29	66.5	9.6	0.01	1.7	0.50	10	2	0.06	89	5	2540	161	1.11E+06
	18/10/2012	18.8	7.77	69.1	9.8	0.01	2.0	0.30	8	2	0.05	89	4	5600	178	1.15E+06
	24/10/2012	19.4	7.64	73.7	9.7	0.004	1.5	0.30	10	1	0.05	107	10	3680	144	9.35E+05
	29/10/2012	20.5	7.73	72.6	9.8	0.004	5.3	0.30	10	1	0.03	107	7	5600	161	9.44E+05
Pr2D	8/10/2012	19.1	7.80	41.2	7.5	0.004	1.7	0.40	15	2	0.03	71	8	2800	**nd**	9.99E+05
	15/10/2012	18.5	8.34	41.9	7.4	0.002	0.8	0.40	8	3	0.05	71	7	6660	88	1.02E+06
	18/10/2012	18.3	7.80	42.3	7.3	0.004	1.7	0.20	4	1	0.08	107	5	8080	167	9.10E+05
	24/10/2012	19.4	7.60	43.6	7.3	0.01	1.0	0.40	7	2	0.05	107	9	4400	125	1.10E+06
	29/10/2012	19.4	7.81	43.8	7.1	0.004	4.4	0.40	10	2	0.02	107	8	3200	95	1.00E+06
Pr3D	8/10/2012	19.3	7.97	57.9	9.8	0.03	1.0	0.4	25	2	0.03	53	8	3760	**nd**	1.15E+06
	15/10/2012	18.8	8.17	61.3	8.8	0.01	1.7	0.3	9	4	0.04	89	6	3460	218	1.11E+06
	18/10/2012	18.3	7.78	61.4	9.7	0.01	1.1	0.4	2	1	0.07	107	4	4400	96	1.06E+06
	24/10/2012	19.5	7.71	69.8	9.9	0.01	1.7	0.4	6	1	0.06	107	7	3680	129	1.09E+06
	29/10/2012	20.0	7.87	67.0	12.3	0.01	5.1	0.2	12	1	0.04	107	7	3640	88	9.92E+05
Pr4D	8/10/2012	19.3	7.76	65.5	9.8	0.001	1.7	0.3	25	3	0.01	71	7	5600	**nd**	1.00E+06
	15/10/2012	19.2	8.40	66.2	7.7	0.005	1.6	0.4	3	2	0.05	107	8	4520	138	1.20E+06
	18/10/2012	18.7	7.97	66.4	8.4	0.002	2.3	0.3	3	0.5	0.09	107	3	8080	105	9.50E+05
	24/10/2012	19.1	7.84	68.3	8.7	0.019	1.0	0.5	12	1	0.05	107	8	6400	265	1.10E+06
	29/10/2012	19.8	7.72	70.8	9.1	0.016	4.1	0.2	15	3	0.04	107	8	6720	195	1.20E+06
Ub1D	8/10/2012	24.4	7.8	653	7.4	0.004	2.9	6.8	15	9	0.6	285	76	32000	**nd**	1.50E+07
	15/10/2012	24.9	7.8	665	6.5	0.01	3.3	1.5	25	8	0.09	356	10	30900	1640	1.90E+07
	18/10/2012	24.3	7.7	660	5.8	0.003	3.1	5	8	12	0.09	267	150	97000	4400	1.85E+07
	24/10/2012	25.6	7.6	643	8.9	0.009	4.9	10	15	2	0.06	356	15	28000	1010	2.20E+07
	29/10/2012	24.3	7.7	639	7.8	0.007	3.9	1.5	15	5	0.03	356	46	21200	2300	2.30E+07
Ub2D	8/10/2012	25.0	7.9	1344	9.4	0.07	7.3	50	50	100	20	356	300	53200	**nd**	2.10E+07
	15/10/2012	26.0	7.9	1298	8.5	0.10	6.3	30	60	250	40	445	300	48400	3900	2.50E+07
	18/10/2012	26.1	7.9	1325	8.7	0.07	7.2	50	50	250	40	445	500	63000	2070	1.90E+07
	24/10/2012	26.0	8.3	1304	8.8	2.51	8.1	10	60	250	40	356	500	69000	5600	2.20E+07
	29/10/2012	27.2	8.0	1417	7.4	0.03	7.9	25	50	250	40	445	500	50000	7000	2.30E+07

Ub3D	8/10/2012	23.9	7.4	636	6.2	0.003	2.4	5	20	5	0.5	178	90	72400	**nd**	2.00E+07
	15/10/2012	31.0	8.0	864	9.5	0.01	3.3	3	10	5	0.1	267	60	79800	33000	2.60E+07
	18/10/2012	30.2	8.2	1443	6.8	0.02	3.7	9	10	3	0.1	267	500	48000	22000	2.50E+07
	24/10/2012	25.8	7.7	1989	3.3	0.01	3.2	25	12	2	0.1	356	500	44000	48000	2.60E+07
	29/10/2012	26.6	7.8	2070	5.9	0.01	3.2	8	15	2	0.1	356	400	26000	24000	3.00E+07
A1D	8/10/2012	26.9	7.0	956	9.7	0.03	5.1	10	25	8	0.8	89	10	109400	**nd**	3.20E+07
	15/10/2012	27.8	7.9	306	9.8	0.29	4.1	25	70	6	0.9	445	30	112000	2100	2.40E+07
	18/10/2012	28.0	7.8	699	4.9	0.04	4.1	10	8	5	0.1	178	29	604000	4000	2.50E+07
	24/10/2012	28.8	8.0	157	7.6	0.03	7.4	9	25	3	0.06	178	15	450000	1200	2.30E+07
	29/10/2012	28.5	7.8	338	9.8	0.13	7.0	12	10	1.5	0.03	107	22	204000	4000	1.80E+07
A2D	8/10/2012	25.4	7.4	308	5.5	0.10	8.4	25	20	6	0.5	89	12	856000	**nd**	1.40E+08
	15/10/2012	29.9	7.4	305	7.6	0.31	6.6	50	50	9	0.8	267	20	972000	4340	1.80E+08
	18/10/2012	32.3	7.1	1067	0.5	0.36	7.1	70	50	8	1.5	267	300	920000	2600	1.77E+08
	24/10/2012	29.9	7.2	754	6.8	0.18	7.3	100	90	2	0.07	178	20	968000	3400	1.60E+08
	29/10/2012	29.2	7.1	851	4.1	0.17	9.2	100	70	1	0.06	178	20	252000	3920	1.70E+08

Appendix 1B: Physico-chemical and microbiological data set for wet season. Abbreviations as in appendix 1A.

Site Id	Sampling dates	Temp °C	pH	Cond µS	DO mg/l	TSS mg/l	BOD mg/L	NH_4^+ mg/l	PO_4^{3-} mg/l	NO_3^- mg/l	NO2 mg/l	Hardness mg/l	Chloride mg/l	HPC cfu/ml	TC cfu/ml	BA cells/ml
Pr1W	08.05.2013	17.4	7.10	40.5	8.6	0.13	2.2	0.30	10	5	0.04	107	25	3400	200	1.01E+06
	18.05.2013	18.0	7.00	46.0	8.7	0.12	0.4	0.30	5	7	0.08	107	12	5400	218	1.76E+06
	29.05.2013	19.0	6.92	48.7	9.8	0.31	0.6	0.25	4	1	0.08	107	80	2700	260	1.80E+06
	05.06.2013	17.2	6.96	50.1	10.2	0.65	1.9	0.20	3	3	0.05	89	70	3520	266	1.84E+06
	12.06.2013	15.6	7.01	52.6	12.6	0.32	1.1	0.40	6	5	0.04	89	20	4800	210	2.00E+06
Pr2W	08.05.2013	16.8	6.80	29.5	8.4	0.12	1.8	0.40	8	6	0.05	89	30	3000	230	1.50E+06
	18.05.2013	17.4	7.10	32.5	7.1	0.41	1.1	0.25	6	6	0.07	89	14	1220	200	1.47E+06
	29.05.2013	18.5	6.80	33.8	8.9	0.24	1.5	0.40	3	2	0.10	107	20	2720	190	1.40E+06
	05.06.2013	16.8	6.68	35.5	9.7	0.26	1.3	0.30	3	2	0.05	71	80	4000	170	1.75E+06
	12.06.2013	15.6	6.69	36.2	10.8	0.26	0.9	0.30	5	3	0.06	89	30	9800	220	1.53E+06
Pr4W	08.05.2013	17.9	7.33	37.3	9.6	0.2	2.0	0.20	5	3	0.03	71	35	2600	230	1.20E+06
	18.05.2013	17.7	7.10	40.5	8.1	0.3	0.6	0.40	6	5	0.09	89	16	1300	260	1.16E+06
	29.05.2013	18.7	7.01	43.6	9.0	0.5	0.7	0.15	5	1	0.09	89	50	3600	190	1.30E+06
	05.06.2013	16.8	7.01	45.1	10.6	0.2	0.9	0.25	3	4	0.04	107	50	4600	250	1.35E+06
	12.06.2013	15.8	6.95	48.8	11.2	0.1	0.9	0.35	10	5	0.04	107	60	2700	200	1.40E+06
Pr4W	08.05.2013	19.5	7.74	52.9	10.6	0.1	0.7	0.30	4	5	0.05	107	28	5500	350	1.11E+06
	18.05.2013	19.1	7.19	55.6	7.1	0.6	1.3	0.35	10	4	0.07	107	18	1600	300	1.40E+06
	29.05.2013	19.3	7.12	58.5	8.6	0.5	1.6	0.20	9	2	0.10	107	50	3000	240	1.20E+06
	05.06.2013	17.8	7.10	58.3	9.6	0.7	0.9	0.20	3	3	0.08	107	70	3800	260	1.43E+06
	12.06.2013	17.1	7.04	59.4	13.7	0.5	1.7	0.40	4	3	0.05	89	40	3600	230	1.38E+06
Ub1W	08.05.2013	21.5	7.5	108	12.3	0.2	2.2	7.0	10	6	0.27	107	230	65000	2340	3.50E+07
	18.05.2013	21.3	7.1	157	14.1	0.7	1.1	7.0	5	6	0.19	107	160	34000	9100	3.70E+07
	29.05.2013	22.4	7.2	197	12.3	0.4	2.1	0.4	5	5	0.10	178	280	46000	2300	4.30E+07
	05.06.2013	20.6	7.2	221	16.7	0.5	3.1	5.0	5	8	0.10	178	220	33000	8400	3.50E+07
	12.06.2013	19.7	7.2	256	15.7	0.1	3.2	4.0	10	8	0.60	178	180	50000	8300	3.40E+07
Ub2W	08.05.2013	26.6	7.7	1125	8.6	0.2	4.8	50	18	9	0.3	356	450	560000	42200	6.90E+08
	18.05.2013	26.9	7.9	1172	12.6	0.7	3.4	60	25	15	0.3	356	380	420000	19000	6.40E+08
	29.05.2013	26.4	7.6	1066	17.1	1.3	5.4	50	10	8	0.3	356	480	585000	12000	6.50E+08
	05.06.2013	24.7	7.7	1312	11.4	0.3	8.1	80	10	10	0.3	267	450	300000	14000	5.70E+08
	12.06.2013	23.6	7.7	1353	12.6	1.0	7.8	90	25	25	3.0	356	500	116000	53000	5.50E+08
Ub3W	08.05.2013	25.6	7.4	162	7.1	0.20	2.7	5	5	7	0.3	178	350	610000	70000	3.90E+07
	18.05.2013	24.7	7.2	167	9.2	0.98	2.1	5	5	7	0.2	107	170	290000	21000	3.94E+07
	29.05.2013	24.5	7.1	186	8.4	0.06	1.9	1	3	7	0.1	178	200	460000	20000	4.00E+07

	05.06.2013	22.8	6.8	212	9.8	0.82	2.4	1	3	4	0.1	178	180	700000	20400	4.34E+07
	12.06.2013	20.8	7.0	214	9.1	0.50	2.1	2	9	6	0.4	107	220	120000	25000	4.50E+07
A1W	08.05.2013	24.5	8.1	845	5.9	0.33	5.4	12	10	7	0.3	267	210	210000	2200	4.30E+08
	18.05.2013	23.6	7.9	954	13.9	8.60	3.7	25	50	15	0.5	356	250	250000	3160	4.70E+08
	29.05.2013	25.1	7.9	979	15.2	0.57	7.0	10	10	5	0.1	356	300	300000	5500	4.10E+08
	05.06.2013	21.9	7.6	719	8.0	0.55	7.2	7	4	8	0.3	267	310	240000	6400	3.90E+08
	12.06.2013	21.3	7.3	909	5.1	0.44	3.2	8	15	25	0.9	267	320	840000	2800	3.10E+08
A2W	08.05.2013	31.4	6.9	289	0.9	1.4	5.3	70	50	10	0.4	267	510	420000	6200	4.80E+08
	18.05.2013	29.9	6.7	265	0.5	2.2	5.2	70	25	10	0.3	445	420	270000	1180	4.78E+08
	29.05.2013	26.4	7.0	384	9.6	6.3	6.3	100	25	8	0.1	356	280	170000	8400	4.60E+08
	05.06.2013	25.2	6.8	489	1.2	1.2	6.8	50	20	9	0.2	267	500	208000	2600	5.20E+08
	12.06.2013	24.6	6.8	273	5.2	4.3	8.5	10	25	10	0.1	178	290	500000	9300	5.50E+08

Appendix 1C: Chlorophyll-*a* and *Clostridium perfringens* data sets for dry and wet seasons.

Site Id	Sampling dates	Chlorophyll-*a* µg/l	*C. perfringens* cfu/ml
Pr1D	6/8/2014	0.29	0.5
	11/8/2014	0.30	0.9
	17/8/2014	0.58	0.4
	23/8/2014	0.31	0.2
	27/8/2014	0.32	0.3
Pr2D	6/8/2014	0.30	0.1
	11/8/2014	0.50	0.4
	17/8/2014	0.42	0.2
	23/8/2014	0.32	0.1
	27/8/2014	0.30	0.2
Pr3D	6/8/2014	0.23	0.7
	11/8/2014	0.32	0.6
	17/8/2014	0.34	0.1
	23/8/2014	0.50	0.2
	27/8/2014	0.43	0.1
Pr4D	6/8/2014	0.30	0.1
	11/8/2014	0.35	0.7
	17/8/2014	0.20	1.3
	23/8/2014	0.30	0.5
	27/8/2014	0.26	0.7
Ub1D	6/8/2014	5.50	19.0
	11/8/2014	3.58	16.0
	17/8/2014	3.30	13.0
	23/8/2014	2.10	11.0
	27/8/2014	3.24	9.0
Ub2D	6/8/2014	9.40	49.0
	11/8/2014	9.28	43.0
	17/8/2014	9.48	43.0
	23/8/2014	9.40	35.0
	27/8/2014	7.80	47.0
Ub3D	6/8/2014	7.20	27.0
	11/8/2014	9.24	28.0
	17/8/2014	9.44	26.0

Site ID	Sampling dates	Chlorophyll-*a* µg/l	*C. perfringens* cfu/ml
Pr1W	08.05.2013	0.32	2.0
	18.05.2013	0.32	2.0
	29.05.2013	0.40	6.9
	05.06.2013	0.29	3.0
	12.06.2012	0.30	3.0
Pr2W	08.05.2013	0.48	1.0
	18.05.2013	0.64	3.0
	29.05.2013	0.24	2.0
	05.06.2013	0.32	1.0
	12.06.2012	0.16	2.0
Pr4W	08.05.2013	0.19	5.0
	18.05.2013	0.32	2.0
	29.05.2013	0.38	4.0
	05.06.2013	0.77	3.0
	12.06.2012	0.16	3.0
Pr4W	08.05.2013	0.32	2.0
	18.05.2013	0.33	5.4
	29.05.2013	0.32	5.0
	05.06.2013	0.16	3.0
	12.06.2012	0.20	2.0
Ub1W	08.05.2013	2.56	50
	18.05.2013	3.52	16
	29.05.2013	3.20	110
	05.06.2013	2.24	21
	12.06.2012	3.20	50
Ub2W	08.05.2013	5.49	180
	18.05.2013	6.24	200
	29.05.2013	7.68	550
	05.06.2013	5.28	97
	12.06.2012	9.60	102
Ub3W	08.05.2013	1.60	26
	18.05.2013	1.70	100
	29.05.2013	1.60	31

	23/8/2014	8.50	30.0
	27/8/2014	8.33	24.0
A1D	6/8/2014	1.80	59.0
	11/8/2014	1.10	110.0
	17/8/2014	2.80	126.0
	23/8/2014	4.00	65.0
	27/8/2014	2.96	45.0
A2D	6/8/2014	5.00	90.0
	11/8/2014	2.00	113.0
	17/8/2014	7.70	130.0
	23/8/2014	7.50	13.0
	27/8/2014	7.52	26.0

	05.06.2013	1.67	27
	12.06.2012	1.60	50
A1W	08.05.2013	1.60	50
	18.05.2013	1.92	165
	29.05.2013	2.40	260
	05.06.2013	4.64	51
	12.06.2012	2.72	58
A2W	08.05.2013	3.48	570
	18.05.2013	3.04	900
	29.05.2013	6.40	32
	05.06.2013	7.40	125
	12.06.2012	7.70	260

Appendix 2A. Nucleotide sequences of primers, targeting 16S rRNA gene, used in the first amplification step (Target specific) of the library preparation for Illumina MiSeq sequencing. Italic lowercase are binding sites for the Illumina sequencing primers. Bold uppercase letters highlight the barcode/index sequence and italic uppercase highlight the linker which links the barcode with the 16S rRNA complementary primer sequence.

Forward Primers	Sequence (5' to 3')
F1	*acactctttccctacacgacgctcttccgatct* **AATGGT** *CA cagcagccgcggtaatac*
F2	*acactctttccctacacgacgctcttccgatct* **ATTCTC** *CA cagcagccgcggtaatac*
F3	*acactctttccctacacgacgctcttccgatct* **ATACCT** *CA cagcagccgcggtaatac*
F4	*acactctttccctacacgacgctcttccgatct* **AATCCA** *CA cagcagccgcggtaatac*
F5	*acactctttccctacacgacgctcttccgatct* **ATTGAG** *CA cagcagccgcggtaatac*
F6	*acactctttccctacacgacgctcttccgatct* **CCTTGA** *CA cagcagccgcggtaatac*
F7	*acactctttccctacacgacgctcttccgatct* **CCGTAG** *CA cagcagccgcggtaatac*
F8	*acactctttccctacacgacgctcttccgatct* **CGGAAC** *CA cagcagccgcggtaatac*
F9	*acactctttccctacacgacgctcttccgatct* **CGATTA** *CA cagcagccgcggtaatac*
F10	*acactctttccctacacgacgctcttccgatct* **CACATA** *CA cagcagccgcggtaatac*
F11	*acactctttccctacacgacgctcttccgatct* **GAATCT** *CA cagcagccgcggtaatac*
F12	*acactctttccctacacgacgctcttccgatct* **GATAAG** *CA cagcagccgcggtaatac*
F13	*acactctttccctacacgacgctcttccgatct* **GGATGC** *CA cagcagccgcggtaatac*
F14	*acactctttccctacacgacgctcttccgatct* **GAACGG** *CA cagcagccgcggtaatac*
F15	*acactctttccctacacgacgctcttccgatct* **GGACTT** *CA cagcagccgcggtaatac*
F16	*acactctttccctacacgacgctcttccgatct* **TGAGGA** *CA cagcagccgcggtaatac*
F17	*acactctttccctacacgacgctcttccgatct* **TACCCA** *CA cagcagccgcggtaatac*
F18	*acactctttccctacacgacgctcttccgatct* **TTCAAC** *CA cagcagccgcggtaatac*
F19	*acactctttccctacacgacgctcttccgatct* **TCATGT** *CA cagcagccgcggtaatac*
F20	*acactctttccctacacgacgctcttccgatct* **TCGCTT** *CA cagcagccgcggtaatac*

Reverse Primer	Sequence (5' to 3')
R0	*gtgactggagttcagacgtgtgctcttccgatct* ccgtcaattcctttgagttt

Appendix 2B. Nucleotide sequences of primers, targeting 16S rRNA gene, used in the second amplification step (Multiplexing) of the library preparation for Illumina MiSeq sequencing. Underlined lowercase letters are binding sites for the Illumina's flow cell and italic lowercase are binding sites for the Illumina sequencing primers. Bold uppercase letters highlight the index sequence.

Reverse Primers	Sequence (5' to 3')
IDX_R1	caagcagaagacggcatacgagat **CGTGAT** *gtgactggagttcagacgtgtgctcttccgatct*
IDX_R2	caagcagaagacggcatacgagat **ACATCG** *gtgactggagttcagacgtgtgctcttccgatct*
IDX_R3	caagcagaagacggcatacgagat **GCCTAA** *gtgactggagttcagacgtgtgctcttccgatct*
IDX_R4	caagcagaagacggcatacgagat **TGGTCA** *gtgactggagttcagacgtgtgctcttccgatct*
IDX_R5	caagcagaagacggcatacgagat **CACTGT** *gtgactggagttcagacgtgtgctcttccgatct*
IDX_R6	caagcagaagacggcatacgagat **ATTGGC** *gtgactggagttcagacgtgtgctcttccgatct*
IDX_R7	caagcagaagacggcatacgagat **GATCTG** *gtgactggagttcagacgtgtgctcttccgatct*
IDX_R8	caagcagaagacggcatacgagat **TCAAGT** *gtgactggagttcagacgtgtgctcttccgatct*
IDX_R9	caagcagaagacggcatacgagat **CTGATC** *gtgactggagttcagacgtgtgctcttccgatct*
IDX_R10	caagcagaagacggcatacgagat **AAGCTA** *gtgactggagttcagacgtgtgctcttccgatct*
IDX_R11	caagcagaagacggcatacgagat **GTAGGC** *gtgactggagttcagacgtgtgctcttccgatct*
IDX_R12	caagcagaagacggcatacgagat **TACAAG** *gtgactggagttcagacgtgtgctcttccgatct*

Forward primer	
Illu_Mplex	aatgatacggcgaccaccgagatct *acactctttccctacacgacgctcttccgatct*

Curriculum Vitae

Personal details

Surname	Mushi
Other names	Douglas Werasimbo
Date of Birth	26 February 1977
Nationality	Tanzanian
Permanent address	P.O. Box 3038, Morogoro, Tanzania
E-mail	douglas@suanet.ac.tz/mushidouglas@yahoo.com

Educational background

2012 - 2015	PhD in Science, Technical University Braunschweig (TUB), Germany
2004 - 2005	MSc. Limnology and wetland ecosystems, UNESCO Institute of Water Education, Delft, The Netherlands
1999 - 2002	BSc. Microbiology and Chemistry, University of Dar es Salaam, Tanzania

Professional experience

2012 - 2015	PhD student in the microbial diagnostic (MIDI) Research group at the Helmholtz Centre for Infection Research, Germany
2007 - present	Lecturer, Biological Sciences Department, Sokoine University of Agriculture

International conference attended:

Mushi D, Pereira RPA, Höfle MG and Brettar I: Bacterial communities and their pathogenic potential along a pollution gradient in a tropical watershed in Tanzania. (Poster) PS-S1.12. *13th SAME conference on aquatic microbial ecology, EMBO, 8-13 September 2013, Stresa, Italy.*

Mushi D, Pereira RPA, Höfle MG and Brettar I: Deep sequencing of microbial communities in tropical drinking water treatment plant reveals significant amplification of rare taxa. (Talk) *18th International symposium on health related water microbiology*, *WaterMicro2015*, *13-19 September, 2015, Lisbon, Portugal*.